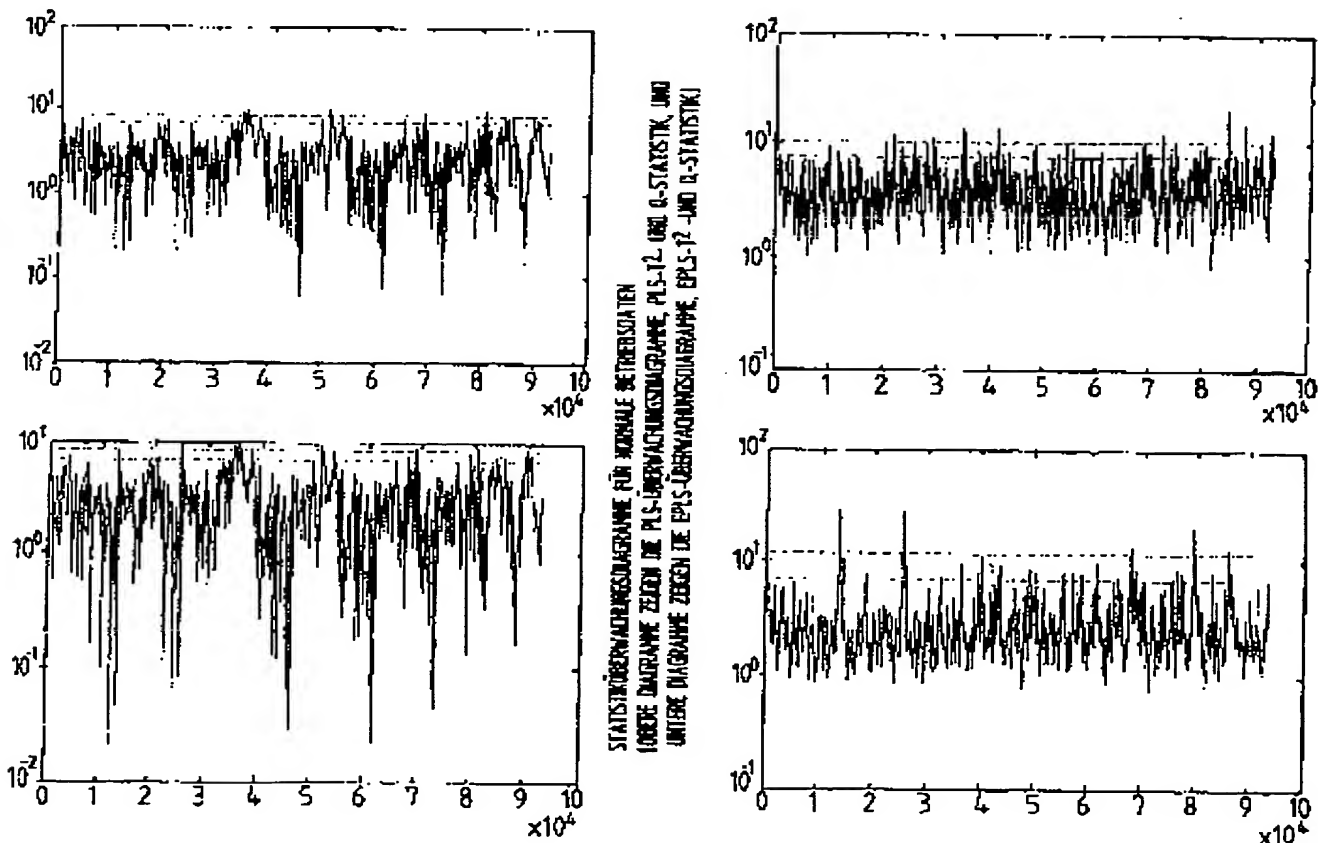


AN: PAT 2002-098793  
 TI: Configuring a multidimensional monitor for statistical processes, involves generating second generalized result vector representing prediction error of a PLS model and residues of a predictor matrix  
 PN: DE10114211-A1  
 PD: 15.11.2001  
 AB: NOVELTY - The method involves generating a first generalized result vector which describes any significant variation of the process, including variations of the predictor and response variables. A second generalized result vector is then generated to represent the prediction error of a PLS model and residues of a predictor matrix. DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for a multidimensional monitor.; USE - Configuring a multidimensional statistical monitor. ADVANTAGE - Represent variations of the predictor and response variables together with their residues. DESCRIPTION OF DRAWING(S) - The figure shows the monitoring diagram for normal operating data. (Drawing includes non-English language text)  
 PA: (CHEN/) CHEN Q; (KRUG/) KRUGER U; (SAND/) SANDOZ D J; (SIMS-) SIMSCI LTD;  
 IN: CHEN Q; KRUGER U; SANDOZ D J;  
 FA: DE10114211-A1 15.11.2001; US2002072882-A1 13.06.2002; GB2365131-A 13.02.2002;  
 CO: DE; GB; US;  
 IC: G01D-003/08; G06F-017/10; G06F-017/18; G08B-029/00;  
 MC: T01-J03; T01-J07A3; T06-A07B; T06-A08;  
 DC: T01; T06;  
 FN: 2002098793.gif  
 PR: GB0007063 23.03.2000;  
 FP: 15.11.2001  
 UP: 08.07.2002





19 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

12 Off nl gungsschrift  
10 DE 101 14 211 A 1

51 Int. Cl.<sup>7</sup>:  
G 06 F 17/18

21 Aktenzeichen: 101 14 211.0  
22 Anmeldetag: 23. 3. 2001  
43 Offenlegungstag: 15. 11. 2001

DE 101 14 211 A 1

30 Unionspriorität:  
0007063. 1 23. 03. 2000 GB  
71 Anmelder:  
Simsci Ltd. Invensys House, London, GB  
74 Vertreter:  
HOFFMANN · EITLE, 81925 München

72 Erfinder:  
Kruger, Uwe, Belfast, IE; Chen, Qian, Prof., Nanjing, CN; Sandoz, David James, Gwynedd, N Wales, GB

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

54 Mehrdimensionale Monitore für statistische Prozesse

57 Es wird eine erweiterte Vorgehensweise mit partiellen kleinsten Fehlerquadraten (EPLS) für die Zustandsüberwachung industrieller Prozesse beschrieben. Diese EPLS-Vorgehensweise stellt zwei statistische Überwachungsdiagramme zur Verfügung, um anomales Prozessverhalten festzustellen, und ebenfalls Beitragsdiagramme, um dieses Verhalten zu diagnostizieren. Es wird eine theoretische Analyse der EPLS-Überwachungsdiagramme zur Verfügung gestellt, zusammen mit zwei Anwendungsstudien, um zu zeigen, dass die EPLS-Vorgehensweise entweder empfindlicher ist oder eine einfachere Interpretation ermöglicht als herkömmliche PLS. Verallgemeinerte Ergebnisse werden durch Konstruktion einer vergrößerten Matrix folgender Form berechnet:

$$Z = [Y; X],$$

wobei X die Prädiktormatrix und Y die Responsmatrix ist, und es wird eine Ergebnismatrix  $T_n = T_n - E_n^*$  konstruiert, bei welcher  $T_n^*$  und  $E_n^*$  folgende allgemeine Form aufweisen:

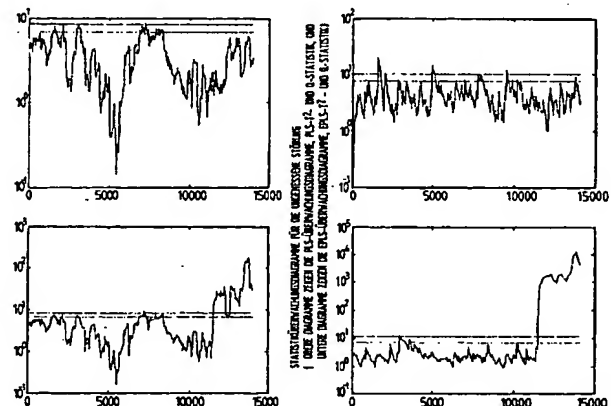
$$T_n^* = [Y; X][B_{n,s}^{(n)}; I]^t R_n$$

$$E_n^* = [E_n; F_n][B_{n,s}^{(n)}; I]^t R_n$$

wobei die Spalten der Matrix  $T_n^*$  die verallgemeinerten t-Ergebnisse zur Verfügung stellen und die Spalten der Matrix  $E_n^*$  die verallgemeinerten restlichen Ergebnisse, wobei eine Einheitsmatrix  $M \times M$  bezeichnet und

$$B_{PLS}^{(n)}$$

die PLS-Regressionsmatrix ist.



DE 101 14 211 A 1

## Beschreibung

[0001] Die vorliegende Erfindung betrifft mehrdimensionale Monitore für statistische Prozesse. Der Begriff "Prozess" soll als allgemeiner Begriff der Regeltheorie verstanden werden und allgemein gesteuerte bzw. geregelte Geräte, Anlagen und derartige Systeme umfassen.

## 1. EINFÜHRUNG

## 1.1 Hintergrund der Erfindung

[0002] Die Erfassung und Diagnose anomaler Situationen im Betrieb industrieller Prozesse stellt ein Problem mit beträchtlicher Herausforderung dar, das sowohl im akademischen als auch industriellen Bereich große Beachtung findet. [Nimmo, 1995] erläuterte, dass allein die petrochemische Industrie in den USA bis zu 10 Milliarden Dollar pro Jahr einsparen könnte, wenn anomale Situationen erfasst, diagnostiziert und entsprechend behandelt werden könnten. Die Folgen der Tatsache, derartige Ereignisse nicht feststellen zu können, können von erhöhten Betriebskosten beim Ablauf eines Prozesses bis zu Produktionsverlusten führen, infolge eines desaströsen Ausfalls der gesamten Anlage.

[0003] Die Aufgabe der Erfassung und Diagnose industrieller Prozesse, kontinuierlich oder postenweise, ist schwierig. Dies liegt daran, dass bei industriellen Prozessen häufig eine große Anzahl an Prozessvariablen vorhanden sind, beispielsweise Temperaturen, Drücke, Flussraten, Zusammensetzungen usw., die regelmäßig bis zu einigen tausend Malen pro Tag aufgezeichnet werden [Piovose, 1991] [Kosanovich, 1992]. Diese sehr große Datenmenge ist schwierig, einfach durch Beobachtung zu analysieren und zu interpretieren. Weiterhin ist es häufig der Fall, dass die Prozessvariablen hochkorreliert sind [MacGregor, 1991] und daher die Anzahl an Freiheitsgraden innerhalb des Prozesses erheblich kleiner ist als die Anzahl beobachteter Prozessvariablen. Dies macht es selbst für einen erfahrenen Bediener schwierig, die Wechselwirkung von Ursache und Wirkung durch Beobachtung zu interpretieren. Allerdings sind in den aufgezeichneten Daten die Grundlagen zur Veranschaulichung des momentanen Zustands der Prozessoperation verborgen. Die Schwierigkeit besteht darin, diese Grundlagen aus den Daten abzu ziehen.

[0004] Um dieses Thema der Erfassung und Diagnose anzugehen, wurden Vorgehensweisen der mehrdimensionalen, statistischen Prozesssteuerung (MSPC) erfolgreich eingesetzt [Kresta, 1991], [MacGregor, 1995], [Kourti, 1995]. Die MSPC-Techniken sind darauf gerichtet, sukzessiv die Anzahl an Variablen zu verringern, die dazu erforderlich sind, signifikante Variationen des Prozesses zu beschreiben. Die aufgezeichneten Daten werden daher in eine Gruppe von weniger Variablen komprimiert, die daher besser handhabbar und interpretierbar sind.

[0005] Eine derartige MSPC-Vorgehensweise sind die partiellen kleinsten Quadrate (PLS), die auf der Pionierarbeit von H. Wold Mitte der Sechziger beruhen [Gelade, 1988]. Die ersten Veröffentlichungen von PLS wurden 1966 vorgestellt [Wold, 1966a; 1966b]. Das PLS-Verfahren identifiziert eine parametrische Regressionsmatrix auf der Grundlage von Prädiktor- und Respons-Matrizen, die aus Referenzdaten des Prozesses aufgebaut werden. Die Prädiktormatrix besteht aus den Signalen der manipulierten und gemessenen Störungs- oder Ursachenvariablen des Prozesses (Prädiktorvariable), wogegen die Responsmatrix aus dem Gesteuerten oder Ergebnisvariablen des Prozesses (Responsvariablen) besteht. Der PLS-Algorithmus zerlegt die Prädiktormatrizen und Responsmatrizen in Komponentenmatrizen des Ranges 1. Jede Komponentenmatrix besteht aus einem Vektorprodukt, bei welchem ein Vektor die Variation (Ergebnisvektor) und der andere den Beitrag (Belastungsvektor) des Ergebnisvektors entweder zur Prädiktormatrix oder zur Responsmatrix beschreibt. Die Zerlegung ist eine Iteration, bei welcher ein Paar von Komponentenmatrizen (eine für den Prädiktor und eine für die Respons) in jedem Iterationsschritt berechnet wird. Die Regressionsmatrix wird bei jedem Iterationsschritt als Ergebnis dieser Zerlegung aktualisiert. Die Datenreduktion wird dadurch erzielt, dass die Variation der Prädiktor- und Responsvariablen bis zur kleinsten Anzahl an Ergebnisvektoren herunterkomprimiert wird, die wirksam das Prozessverhalten beschreiben können. Die Auswahl der Anzahl an Komponentenmatrizen, die beibehalten werden müssen, stellt einen Kompromiss zwischen dem Maximieren der Variation, die in den Prädiktor- und Responsmatrizen erläutert wird, und dem Minimieren der Anzahl an Komponentenmatrizen dar. Kreuzüberprüfung [Wold, 1978] wird am häufigsten dazu verwendet, die Anzahl zurückzuhaltender Komponentenmatrizen zu definieren, vgl. beispielsweise [MacGregor, 1991, 1995], [Morud, 1996].

[0006] [MacGregor, 1995] und [Wise, 1996] gaben an, dass die PLS-Zerlegung der Prädiktormatrix für die Zustandsüberwachung kontinuierlicher, industrieller Prozesse verwendet werden kann. Weiterhin betonen sie, dass diese Zerlegung ähnlich einer Hauptkomponentenanalyse (PCA) der Prädiktormatrix ist. Die PLS-Zerlegung der Prädiktormatrix gestattet die Berechnung von zwei Statistiken. Die erste Statistik (die T-Quadratstatistik) beschreibt die Variation der Prädiktormatrix, die signifikant zur Vorhersage der Responsvariablen ist. Im Gegensatz hierzu entspricht die zweite Statistik (die Q-Statistik) der Variation bei der Prädiktormatrix, die insignifikant zur Vorhersage der Responsvariablen ist. Beide Statistiken können in statistischen Überwachungsdiagrammen in Abhängigkeit von der Zeit aufgetragen werden. Diese Vorgehensweise wird nachstehend als "Vorgehensweise I" bezeichnet.

[0007] Eine weitere Vorgehensweise zur Nutzung von PLS als ein Zustandsüberwachungswerkzeug wird beispielsweise in [MacGregor, 1991] und [Kresta, 1991] diskutiert. Bei dieser Vorgehensweise werden verschiedene statistische Plots dazu verwendet, ein anomales Prozessverhalten festzustellen und zu diagnostizieren. Diese Plots sind:

- x-y-Plots des quadrierten Vorhersagefehlers der Responsvariablen in Abhängigkeit von den Ergebniswerten jedes Ergebnisvektors, der die Prädiktorvariablen repräsentiert (Überwachungsdiagramme),
- Plots jeder Kombination der beiden Ergebnisvektoren, welche die Prädiktorvariablen repräsentieren (Streuungs-Plots), und
- Plots des quadrierten Vorhersagefehlers, entweder der Prädiktor- oder Responsvariablen, in Abhängigkeit von der Zeit (SPE-Diagramme).

[0008] Diese Vorgehensweise wird in der nachstehenden Beschreibung als "Vorgehensweise II" bezeichnet.

## 1.2 Zusammenfassung der Erfindung

[0009] In der folgenden Beschreibung wird eine Ausweitung des Standard-PLS-Algorithmus, nachstehend als "erweiterter PLS" oder "EPLS" bezeichnet, für kontinuierliche Prozesse erläutert. Diese Erweiterung führt zu der Bestimmung zweier neuer PLS-Ergebnisse auf der Grundlage der Ergebnisvektoren der Prädiktormatrix. Die neuen PLS-Ergebnisvektoren werden als verallgemeinerte Ergebnisvektoren bezeichnet. Der erste verallgemeinerte Ergebnisvektor beschreibt eine signifikante Variation des Prozesses einschließlich der Prädiktor- und Responsvariablen. Der zweite verallgemeinerte Ergebnisvektor repräsentiert den Vorhersagefehler des PLS-Modells und Residuen der Prädiktormatrix. Die EPLS-Vorgehensweise führt zu Überwachungsdiagrammen für T-Quadrat und Q, die ähnlich jenen sind, die bei PCA erhalten werden, wenn sowohl Prädiktor- als auch Responsvariablen mittels PCA analysiert werden. Dies unterscheidet sich von der üblichen PLS-Vorgehensweise, welche nur die Prädiktorvariablen analysiert und daher keinen Aufschluss über das Verhalten der Responsevariablen ermöglicht, es sei denn, in dem Prozess wäre eine Rückkopplung vorhanden.

[0010] Der Vorteil der EPLS-Überwachungsdiagramme besteht daher darin, dass sie eine Variation der Prädiktor- und Responsvariablen zusammen mit ihren Residuen repräsentieren. Dies verbessert die Überwachungsdiagramme der Vorgehensweise I, welche nur die Variation und Residuen der Prädiktorvariablen beschreiben. Im Gegensatz zur Vorgehensweise II, stellt EPLS die Fähigkeit zur Verfügung, den Prozess auf nur zwei Diagrammen zu überwachen, im Gegensatz zu bisher, als die Anzahl an Diagrammen von der Anzahl an Komponentenmatrizen abhängig.

[0011] Gemäß einer Zielrichtung der vorliegenden Erfindung umfasst ein Verfahren für die Konstruktion und das Konfigurieren eines mehrdimensionalen Monitors für statistische Prozesse durch eine Vorgehensweise mit partiellen kleinsten Quadraten die Konstruktion der Prädiktor- und Responsmatrizen des Prozesses aus Referenzdaten, wobei die Prädiktormatrix aus Signalen der manipulierten und gemessenen Störungs- oder Ursachenvariablen des Prozesses (Prädiktorvariablen) besteht, und die Responsmatrix aus den gesteuerten oder Auswirkungsvariablen des Prozesses (Responsvariablen) besteht, Zerlegen der Prädiktor- und Responsmatrizen in Komponentenmatrizen des Rangs 1, wobei jede der Komponentenmatrizen aus einem Vektorprodukt besteht, bei welchem ein Vektor (der Ergebnisvektor) die Variation beschreibt, und der andere (der Belastungsvektor) den Beitrag des Ergebnisvektors zu der Prädiktor- oder Responsmatrix beschreibt, und die Zerlegung durch die Erzeugung einer parametrischen Regressionsmatrix durchgeführt wird, auf der Grundlage von Iterationen der Zerlegung der Prädiktor- und Responsmatrizen, und ist gekennzeichnet durch die Erzeugung eines ersten verallgemeinerten Ergebnisvektors, der jede signifikante Variation des Prozesses beschreibt, einschließlich Variationen der Prädiktor- und Responsvariablen, sowie eines zweiten verallgemeinerten Ergebnisvektors, welcher den Vorhersagefehler des Modells der partiellen kleinsten Quadrate und Residuen der Prädiktormatrix repräsentiert.

[0012] Vorzugsweise werden die verallgemeinerten Ergebnisse durch Konstruktion einer vergrößerten Matrix berechnet, die hiermit mit Z bezeichnet ist und folgende Form aufweist:

$$Z = [Y : X],$$

wobei X die Prädiktormatrix und Y die Responsmatrix ist, und durch Konstruktion einer Ergebnismatrix  $T_n = T_n^* - E_n^*$ , wobei  $T_n^*$  und  $E_n^*$  folgende allgemeine Form aufweist:

$$T_n^* = [Y : X] [B_{PLS}^{(n)} : I]^T R_n$$

$$E_n^* = [E_n : F_n] [B_{PLS}^{(n)} : I]^T R_n$$

wobei die Spalten der Matrix  $T_n^*$  die verallgemeinerten t-Ergebnisse angeben, und die Spalten der Matrix  $E_n^*$  die verallgemeinerten, restlichen Ergebnisse, wobei I eine Einheitsmatrix  $M \times M$  bezeichnet, und  $B_{PLS}^{(n)}$  die PLS-Regressionsmatrix ist.

[0013] Gemäß einer zweiten Zielrichtung der Erfindung stellen wir einen mehrdimensionalen Monitor für statistische Prozesse zur Verfügung, der gemäß der ersten Zielrichtung der Erfindung konstruiert und konfiguriert ist, und der so ausgebildet ist, dass er anomales Prozessverhalten durch Analyse der Residuen der Responsvariablen identifiziert.

[0014] Gemäß einer dritten Zielrichtung der Erfindung stellen wir ein Verfahren zur Überwachung eines Prozesses zur Verfügung, welches das Konfigurieren eines mehrdimensionalen Monitors für statistische Prozesse mit dem Verfahren gemäß der ersten Zielrichtung der Erfindung umfasst, und zum Identifizieren eines anomalen Prozessverhaltens, zumindest teilweise durch Analyse der Residuen der Responsvariablen.

[0015] Die Erfindung wird nachstehend, nur beispielhaft, unter Bezugnahme auf die beigefügten Figuren beschrieben. Es zeigt:

[0016] Fig. 1 eine schematische Darstellung der katalytischen Fluidkrackeinheit,

[0017] Fig. 2 eine schematische Darstellung eines Fließbettreaktors und seiner benachbarten Einheit.

[0018] Figuren für die katalytische Fluidkrackeinheit ohne Regelungsrückkopplung in der Prädiktormatrix:

[0019] Fig. 3 Statistiküberwachungsdiagramme für normale Betriebsdaten (die oberen Diagramme geben die PLS-Überwachungsdiagramme an, PLS- $T^2$ - und Q-Statistiken, und die unteren Diagramme zeigen die EPLS-Überwachungsdiagramme, EPLS- $T^2$ - und Q-Statistiken),

[0020] Fig. 4 statistische Überwachungsdiagramme für die ungemessene Störung (die oberen Diagramme geben die PLS-Überwachungsdiagramme an, PLS- $T^2$ - und Q-Statistiken, und die unteren Diagramme zeigen die EPLS-Überwachungsdiagramme, EPLS- $T^2$ - und Q-Statistiken),

[0021] Fig. 5 Fehlerbeitragsdiagramm für den Zeitpunkt 11460 Minuten. Die  $O_2$ - und die CO-Konzentration in dem

Dichtgasfluss weisen den größten Vorhersagefehler auf,

[0022] Fig. 6 statistische Überwachungsdiagramme für die Änderung des Flusses des regenerierten Katalysators in den Reaktor (die oberen Diagramme zeigen die PLS-Überwachungsdiagramme, PLS-T<sup>2</sup>- und Q-Statistiken, und die unteren Diagramme zeigen die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistiken),

5 [0023] Fig. 7 Fehlerbeitragsdiagramm für die Änderung des Flusses des regenerierten Katalysators zum Reaktor zum Zeitpunkt 19357 Minuten. Der Steigrohr-Katalysator-Pegel und die O<sub>2</sub>-Konzentration in dem Gichtgas werden am stärksten beeinflusst.

[0024] Figuren für die katalytische Fluidkrackeinheit mit Regelungsrückkopplung in der Prädiktormatrix:

10 [0025] Fig. 8 statistische Überwachungsdiagramme für die ungemessene Störung (Verkokungsfaktor); Prädiktorvariablen umfassen das Feuchtgaskompressor-Saugventil (obere Diagramme sind die PLS-Überwachungsdiagramme, PLS-T<sup>2</sup>- und Q-Statistiken, und die unteren Diagramme sind die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistiken), zum Zeitpunkt 19357 Minuten.

15 [0026] Fig. 9 statistische Überwachungsdiagramme für die Änderung des Flusses des regenerierten Katalysators zum Reaktor; Prädiktorvariablen umfassen das Feuchtgaskompressor-Saugventil (obere Diagramme sind die PLS-Überwachungsdiagramme, PLS-T<sup>2</sup>- und Q-Statistiken, und die unteren Diagramme sind die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistiken).

[0027] Figuren des Fließbettreaktionsprozesses:

20 [0028] Fig. 10 Statistiküberwachungsdiagramme für Normalbetriebsdaten (oberes Diagramm ist das PLS-Überwachungsdiagramm, PLS-T<sup>2</sup>- und Q-Statistiken, und die unteren Diagramme sind die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistiken),

[0029] Fig. 11 statistische Überwachungsdiagramme für die ungemessene Störung (oberes Diagramm ist das PLS-Überwachungsdiagramm, PLS-T<sup>2</sup>- und Q-Statistiken, und die unteren Diagramme sind die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistik),

25 [0030] Fig. 12 EC-Diagramme für den Dampfdruckumschlag zu den Zeitpunkten 1500 Minuten (oberer linker Plot), 1501 Minuten (unterer linker Plot) und 1502 Minuten (oberer rechter Plot),

[0031] Fig. 13 statistische Überwachungsdiagramme für ein anomales Verhalten eines der Rohre (oberes Diagramm ist das PLS-Überwachungsdiagramm, PLS-T<sup>2</sup>- und Q-Statistik, untere Diagramme sind die EPLS-Überwachungsdiagramme, EPLS-T<sup>2</sup>- und Q-Statistiken),

30 [0032] Fig. 14 EC-Diagramme für das Fluidisierungsproblem in einem der Rohre zu den Zeitpunkten 436 Minuten (oberes linkes Diagramm), 888 Minuten (oberes rechtes Diagramm), 905 Minuten (unten links) und 910 Minuten (unten rechtes Diagramm).

35 [0033] Um die Nützlichkeit der EPLS-Überwachungsdiagramme zu demonstrieren und einen Vergleich mit der Vorgehensweise I durchzuführen, werden nunmehr zwei Fallbeispiele betrachtet. Für jedes werden zwei typische Bedingungen für anomales Verhalten erzeugt, welche die Auswirkung einer ungemessenen Störung und ebenso eine "interne" Änderung des Prozessverhaltens beschreibt. Die Prozessbeispiele sind die Simulation einer katalytischen Fluidkrackeinheit (FCCU), eingeführt von [McFarlane, 1993], und ein realer industrieller Prozess, der zwei unterschiedliche Lösungsmittel als Ergebnis einer komplizierten chemischen Reaktion erzeugt, die in einem Fließbettreaktor durchgeführt wird.

40 [0034] Die folgende Beschreibung ist folgendermaßen organisiert. Im nachstehenden Abschnitt 2 werden sowohl der Standard-PLS-Algorithmus als auch der neue EPLS-Algorithmus beschrieben und verglichen. Im Abschnitt 3 werden die Zustandsüberwachungsstatistiken eingeführt, welche diesen beiden Vorgehensweisen zugeordnet sind. Abschnitt 4 gibt die Anwendungsfalluntersuchungen an, um Beispiele für die Vorteile von EPLS anzugeben.

## 2. ALGORITHMEN PARTIELLER KLEINSTER QUADRATE

### 2.1 Der Standard-PLS-Algorithmus

45

[0035] Die Standard-PLS-Identifizierungstechnik beruht auf der Zerlegung der Prädiktormatrix  $X_0 \in \mathbb{R}^{K \times N}$  und der Responsmatrix  $Y_0 \in \mathbb{R}^{K \times N}$  in eine Summe von Komponentenmatrizen des Rangs 1 [Gelade, 1986]. Beide Matrizen enthalten K Datenpunkte, die Prädiktormatrix besteht aus M Variablen, und die Responsmatrix aus N Variablen. Beide Matrizen werden gewöhnlich vor der Identifizierungsprozedur mittenzentriert und geeignet skaliert. Die Zerlegung beider Matrizen erfolgt folgendermaßen:

$$\begin{aligned}
 X_0 &= \sum_{i=1}^M \tilde{X}_i = \sum_{i=1}^M t_i p_i^T = T_M \cdot P_M^T \\
 Y_0 &= \sum_{i=1}^M \tilde{Y}_i + E_M = \sum_{i=1}^M \hat{u}_i q_i^T + E_M = \hat{U}_M \cdot Q_M^T + E_M
 \end{aligned}
 \tag{1}$$

55

60 wobei X und Y die Komponentenmatrix der Prädiktor- bzw. Responsmatrix ist. Gemäß Gleichung (1) können die Matrizen des Rangs Eins als Vektorprodukt zwischen  $t_i$  und  $\hat{u}_i$ , definiert als Ergebnisvektoren oder latente Variablen (LVs), und von  $p_i$  und  $q_i$ , definiert als Belastungsvektoren, berechnet werden. M ist gleich der Anzahl an Prädiktorvariablen, und  $E_M$  repräsentiert den Vorhersagefehler des Prozessmodells. Es wird darauf hingewiesen, dass dann, wenn alle Komponentenmatrizen eingeschlossen sind, die Prädiktormatrix gleich den Matrixzerlegungen ist. Wenn nur n Komponentenmatrizen eingeschlossen sind, dann wird Gleichung (1) zu:

65

$$\begin{aligned} X_0 &= \sum_{i=1}^n \tilde{X}_i = \sum_{i=1}^n t_i p_i^T = T_n \cdot P_n^T + F_n \\ Y_0 &= \sum_{i=1}^n \tilde{Y}_i + E_n = \sum_{i=1}^n \hat{u}_i q_i^T + E_n = \hat{U}_n \cdot Q_n^T + E_n \end{aligned} \quad (2)$$

wobei  $F_n$  die Residuen der Prädiktormatrix bezeichnen. Die vorhergesagten u-Ergebnisse  $\hat{u}_i$  können durch die folgende Multiplikation bestimmt werden:

$$\hat{U}_n = [t_1 b_1 : \dots : t_n b_n] = T_n \text{diag}\{b_n\}, \quad (3)$$

wobei  $\text{diag}\{b_n\}$  eine Diagonalmatrix ist, welche die Regressionskoeffizienten  $b_i$  des Ergebnismodells aufeinanderfolgend enthält. Eine theoretische Untersuchung des PLS-Algorithmus findet sich im Anhang 1. Unterschiedliche Vorgehensweisen wurden eingeführt, um die Ergebnis- und Belastungsvektoren zu bestimmen, und dies sind der LSQR-Algorithmus [Manne, 1987], der NIPALS-Algorithmus [Geladi, 1986], der SIMPLS-Algorithmus [de Jong, 1993] und andere. [0036] Infolge der Tatsache, dass industrielle Prozesse häufig stark korrelierte Prozessvariablen aufweisen, können nur einige wenige LVs dazu erforderlich sein, den Hauptanteil der Prozessvariation zu beschreiben. Im Gegensatz hierzu betreffen die verbleibenden Paare der LVs im Wesentlichen Rauschen und nicht-signifikante Variationen bei  $X_0$  und  $Y_0$  [Gelade, 1986], [MacGregor, 1991] und [Wise, 1996]. Um die Anzahl zurückzuhaltender LVs zu bestimmen, wurden Kreuzüberprüfung [Wold, 1978] und die Analyse der Varianz (ANOVA) [Jackson, 1991] diskutiert.

## 2.2 Ableitung des EPLS-Algorithmus und der verallgemeinerten Ergebnisse

[0037] Der EPLS-Algorithmus erzeugt Ergebnisse, welche die Variation der Prädiktor- und Responsvariablen repräsentieren, sowie deren Residuen, und diese werden als verallgemeinerte Ergebnisse bezeichnet. Diese Ergebnisse bilden die Grundlage für eine wirksamere Prozesszustandsüberwachung als bei den vorhandenen Vorgehensweisen, die hauptsächlich auf Ergebnissen beruhen, welche nur die Variation der Prädiktorvariablen beschreiben. Die verallgemeinerten Ergebnisse werden berechnet, nachdem die Gewichtung- und Belastungsmatrizen bestimmt wurden (siehe Anhang 1), und beruhen auf der Vergrößerung der Responsmatrix zu der Prädiktormatrix. Die vergrößerte Matrix wird mit  $Z$  bezeichnet und ist folgendermaßen definiert:

$$Z = [Y : X] \quad (4).$$

[0038] Es wird darauf hingewiesen, dass der Index 0 bei beiden Matrizen weggelassen ist. Dies liegt daran, dass die Ableitung der verallgemeinerten Ergebnisse auf dem Standard-PLS-Algorithmus beruht und der Druckablassvorgang nicht erneut durchgeführt werden muss.

[0039] Aus der Gleichung (4) erhält man durch Subtraktion der vorhergesagten Responsmatrix  $Y_n$  und der rekonstruierten Prädiktormatrix  $X_n$ , wobei eine Anzahl von  $n$  an LVs beibehalten wird, den folgenden Ausdruck:

$$[Y : X] - [\hat{Y}_n : \hat{X}_n] = [E_n : F_n], \quad (5)$$

wobei  $E_n$  der Vorhersagefehler der Responsmatrix ist, und  $F_n$  die Residuen der Prädiktormatrix bezeichnen. Durch Einsetzen der Gleichungen (2) und (3) kann Gleichung (5) folgendermaßen umgeschrieben werden:

$$[Y : X] - T_n [\text{diag}\{b_n\} Q_n^T P_n^T] = [E_n : F_n]. \quad (6)$$

[0040] Wie im Anhang 2 gezeigt, ist das Matrixprodukt  $P_n^T$  mit der PLS-Regressionsmatrix  $B_{PLS}^{(n)}$ , wobei  $n$  LVs zurückgehalten werden, gleich dem Matrixprodukt der Diagonalmatrix  $\text{diag}\{b_n\}$  und  $Q_n^T$ . Das Integrieren dieses Ergebnisses in Gleichung (6) ergibt:

$$[Y : X] - T_n P_n^T [B_{PLS}^{(n)} : \mathcal{I}] = [E_n : F_n], \quad (7)$$

wobei  $\mathcal{I}$  eine Einheitsmatrix  $M \times M$  bezeichnet. Die Ausführung einer Multiplikation von Gleichung (7) mit dem verallgemeinerten Inversen von  $[B_{PLS} : \mathcal{I}]$  ergibt:

$$[Y : X] [B_{PLS}^{(n)} : \mathcal{I}]^\dagger - T_n P_n^T = [E_n : F_n] [B_{PLS}^{(n)} : \mathcal{I}]^\dagger, \quad (8)$$

wobei  $\dagger$  das verallgemeinerte Inverse bezeichnet. Wie im Anhang 3 gezeigt, führt die Multiplikation von Gleichung (8) mit  $R_n$  (siehe Anhang 1) zu einer Formel zur Berechnung der Ergebnisse der Prädiktormatrix  $T_n$ :

$$T_n = [Y : X] [B_{PLS}^{(n)} : \mathcal{I}]^\dagger R_n - [E_n : F_n] [B_{PLS}^{(n)} : \mathcal{I}]^\dagger R_n. \quad (9)$$

[0041] In Gleichung (9) ist die Ergebnismatrix  $T_n$  gleich der Differenz von zwei Matrizen. Die erste Matrix bezieht sich auf die Prädiktor- und die Responsmatrix, und die zweite Matrix hängt von dem Vorhersagefehler der Responsmatrix und den Residuen der Prädiktormatrix ab. Die Matrix  $[E_n : F_n]$  wird als die vergrößerte Residuenmatrix von  $F_n$  be-

zeichnet.

[0042] Definiert man die Matrix

$$\left[ \mathbf{B}_{\text{PLS}}^{(n)} : \mathbf{I} \right]^t \mathbf{R}_n$$

als  $\mathbf{C}_{\text{PLS}}^{(n)}$ , die sich ergebende Matrix  $[\mathbf{Y} : \mathbf{X}] \mathbf{C}_{\text{PLS}}^{(n)}$  als  $\mathbf{T}_n^*$ , und das Matrixprodukt  $[\mathbf{E}_n : \mathbf{F}_n] \mathbf{C}_{\text{PLS}}^{(n)}$  als  $\mathbf{E}_n^*$ , so vereinfacht sich Gleichung (9) folgendermaßen:

$$\mathbf{T}_n \mathbf{E} \mathbf{T}_n^* - \mathbf{E}_n^*. \quad (10)$$

[0043] Die Spalten der Matrix  $\mathbf{T}_n^*$  werden als die verallgemeinerten t-Ergebnisse bezeichnet, während die Spalten der Matrix  $\mathbf{E}_n^*$  als verallgemeinerte Residuenergebnisse bezeichnet werden. Für die Prozesszustandsüberwachung stellt Gleichung (10) Ergebnisse zur Verfügung, welche die Prozessvariation beschreiben, die innerhalb der Absolutwerte der Prädiktor- und der Responsmatrix enthalten sind, sowie die Vorhersagefehlermatrix  $\mathbf{E}_n$ , als auch die Residuen der Prädiktormatrix  $\mathbf{F}_n$ .

[0044] Der nächste Abschnitt beschreibt die Ableitung der Statistiken für  $\mathbf{T}_n^*$  und  $\mathbf{E}_n^*$ , die in Abhängigkeit von der Zeit bei eindimensionalen Überwachungsdiagrammen aufgetragen werden können.

### 2.3 Vergleich mit den Vorgehensweisen I und II für die Prozesszustandsüberwachung

[0045] Wie voranstehend erwähnt, beruhen die vorhandenen Vorgehensweisen zur Prozesszustandsüberwachung hauptsächlich auf den t-Ergebnissen. [MacGregor, 1995] und [Wise, 1996] führten aus, dass bei der Vorgehensweise I die weggefallenen und die beibehaltenen t-Ergebnisse die Grundlage für zwei Überwachungsdiagramme bilden, die in dem nächsten Unterabschnitt diskutiert werden. Obwohl erfolgreiche Anwendungen der Vorgehensweise I diskutiert wurden, beispielsweise [Kourti, 1995], [Wise, 1996] und [Morud, 1996], stellen sie nicht notwendigerweise jede Art eines anomalen Prozessverhaltens fest. Dies gilt insbesondere, wenn:

1. anomales Prozessverhalten hauptsächlich die Responsvariablen beeinflusst, die nicht durch eine geschlossene Regelschleife geregelt werden. In diesem Fall breitet sich das anomale Verhalten nicht bis zu den Prädiktorvariablen mittels Regelungsrückkopplung aus, und bleibt daher unentdeckt. Mit EPLS wird die Variation der Responsvariablen deutlich.

2. Die Responsvariablen sind stark korreliert, jedoch die Prädiktorvariablen nicht. In diesem Fall kann nur ein statistisches Diagramm für die Standard-PLS-Vorgehensweise erhalten werden, das Diagramm für quadriertes T. Bei EPLS bleiben beide Diagramme relevant, unabhängig von der Anzahl beibehaltener LVs.

[0046] Die Vorgehensweise II verlässt sich auf Streuungsplots und x-y-Diagramme von SPE in Abhängigkeit von verschiedenen t-Ergebnissen und SPE-Diagrammen, beispielsweise [MacGregor, 1991] und [Kresta, 1991]. Wenn der Prozess aus einer großen Anzahl stark korrelierter Prozessvariablen besteht, beispielsweise einhundert oder mehr, kann die Anzahl erforderlicher Ergebnisse allerdings immer noch groß sein, damit signifikante Prozessvariationen erfasst werden. Daher kann bei diesen anderen Vorgehensweisen eine große Anzahl an Diagrammen erforderlich sein, und wird die Situation mühsam zu analysieren. Im Gegensatz hierzu sind bei EPLS nur zwei Diagramme erforderlich, unabhängig von den Dimensionen des Problems und von der Anzahl ausgewählter LVs.

[0047] Man könnte auch das SPE-Diagramm von Responsvariablen zusätzlich zur Vorgehensweise I dazu verwenden, die voranstehenden Nachteile zu überwinden. Hierbei würden jedoch zumindest drei Überwachungsdiagramme benötigt, und wird die Variation der Responsvariablen nicht in einem dieser Diagramme akkumuliert. Im Gegensatz hierzu erfordern die verallgemeinerten Ergebnisse nur zwei Überwachungsdiagramme, und erfasst eines der generalisierten Ergebnisse die Variation der Responsvariablen. Schließlich wird darauf hingewiesen, dass die Verwendung der generalisierten Ergebnisse zur Prozesszustandsüberwachung ähnlich der Art und Weise ist, in welcher PCA zur Überwachung industrieller Prozesse verwendet wird [Jackson, 1991].

## 3. STATISTIK DER PLS- UND EPLS-VORGEHENSWEISE

### 3.1 Statistik für PLS

[0048] Bei der Vorgehensweise I können zwei statistische Überwachungsdiagramme erhalten werden, auf der Grundlage der Zerlegung der Prädiktormatrix. Das erste Überwachungsdiagramm steht in Beziehung zu den beibehaltenen t-Ergebnissen und beschreibt einen signifikanten Beitrag zur Vorhersage der Responsmatrix. Das zweite Diagramm ist der Variation der Prädiktormatrix zugeordnet, die von den weggefallenen t-Ergebnissen erfasst wird. Die weggefallenen t-Ergebnisse beschreiben nicht-signifikante und unkorrelierte Beiträge zur Vorhersage der Responsmatrix. Allerdings muss in einem Fall, in welchem sämtliche Prädiktorvariablen signifikant zur Variation der Responsvariablen beitragen, jedes t-Ergebnis beibehalten werden. Daher bleiben keine t-Ergebnisse zur Berechnung des zweiten Überwachungsdiagramms übrig.

[0049] Das erste Überwachungsdiagramm beruht auf einer Statistik, die als PLS- $T^2$ -Statistik bezeichnet wird, und das zweite Überwachungsdiagramm betrifft eine Statistik, die als PLS-Q-Statistik bezeichnet wird. Diese beiden Statistiken sind folgendermaßen definiert:

$$\begin{aligned} {}_{\text{PLS}}T_k^2 &= \sum_{i=1}^n \left( \frac{f_{ki}}{{}_T\sigma_i} \right)^2 \\ {}_{\text{PLS}}Q_k &= \sum_{i=1}^M \left( \frac{f_{ki}}{{}_X\sigma_i} \right)^2 \end{aligned} \quad (11)$$

wobei  ${}_{\text{PLS}}T_k^2$  und  ${}_{\text{PLS}}Q_k$  die PLS- $T^2$ - bzw. die Q-Statistik bezeichnen. Weiterhin bezeichnet  $t_{ki}$  den Wert des i-ten t-Ergebnisses zum Zeitpunkt k, und bezeichnet  ${}_T\sigma_i$  die Standardabweichung des i-ten Ergebnisvektors der Prädiktorvariablen der Bezugsdaten.  $f_{ki}$  repräsentiert die Residuen der i-ten Prädiktorvariablen zum Zeitpunkt k, und  ${}_X\sigma_i$  ist die Standardabweichung der i-ten Residuenvariablen der Bezugsdaten. Die Bezeichnungen  $T^2$  und Q wurden entsprechend der  $T^2$ - und Q-Statistik von Hotelling ausgewählt, die bei PCA verwendet werden. Jede Statistik kann in einem Überwachungsdiagramm in Abhängigkeit von der Zeit aufgetragen werden. Es wird darauf hingewiesen, dass die Normierung von  $t_{ki}$  und  $f_{ki}$  unbedingt erforderlich ist, um eine empfindliche Statistik zur Verfügung zu stellen. Falls dies nicht erfolgt, dominieren die t-Ergebnisse mit einer großen Varianz, normalerweise die wenigen ersten, den Ergebniswert von PLS- $T^2$ , und die Residuen der Prädiktorvariablen, die eine große Varianz aufweisen, verdecken Residuen mit relativ kleiner Varianz. Ein Fehlerzustand, der hauptsächlich die t-Ergebnisse oder Residuen beeinflusst, welche kleine Änderungen aufweisen, kann in diesem Fall unentdeckt bleiben. Darüber hinaus kann in Bezug auf die Summe stochastischer Variablen mit Mittelwert Null und Einheitsvarianz (Chi-Quadrat-Verteilung) die statistische Bestimmung von Schwellenwerten verwendet werden.

[0050] Wenn außergewöhnlich große PLS- $T^2$ -Werte auftreten, dann ist die gesamte Prozessvariation ungewöhnlich groß, verglichen mit den Bezugsdaten des Prozesses. Dies bedeutet, dass das allgemeine Prozessverhalten sich beträchtlich geändert hat, oder dass sich der Prozess zu einem neuen Betriebsbereich bewegt hat. Im Gegensatz hierzu zeigen ungewöhnlich große PLS-Q-Werte an, dass sich die Beziehungen zwischen den Prädiktorvariablen in Bezug auf die Beziehung geändert haben, die bei den Bezugsdaten vorherrscht.

### 3.2 Statistik für EPLS

[0051] Im Vergleich zu den Standard-t-Ergebnissen sind die statistischen Eigenschaften der verallgemeinerten Ergebnisse nachstehend zusammengefasst.

1. Die verallgemeinerten t-Ergebnisse und ebenso die verallgemeinerten restlichen Ergebnisse sind um den Mittelwert zentriert, wenn die Spalten in der Prädiktor- und der Responsmatrix in Bezug auf den Mittelwert vor der PLS-Identifizierung zentriert wurden.
2. Die t-Ergebnisse des Standard-PLS-Algorithmus sind orthogonal [Hoskuldsson, 1988]. Im Gegensatz zu den Standard-t-Ergebnissen sind bei EPLS beide Ergebnistypen nicht orthogonal, unabhängig von der Anzahl zurückgehaltener LVs. Ein Beweis findet sich im Anhang 4.

[0052] Um die verallgemeinerten Ergebnisse für die Prozesszustandsüberwachung zu analysieren, ist es wünschenswert, statistisch unabhängige Ergebnisse zu haben, was Orthogonalität erfordert. Um orthogonale Ergebnisse zu erzielen, kann eine Singulärwertzerlegung (SVD) [Golub, 1996] der verallgemeinerten Ergebnisse eingesetzt werden, woraus sich ergibt:

$$\begin{aligned} T_n^* &= V_n^{(T^*)} \Lambda_n^{(T^*)} W_n^{(T^*)T} \\ E_n^* &= V_n^{(E^*)} \Lambda_n^{(E^*)} W_n^{(E^*)T} \end{aligned} \quad (12)$$

wobei  $V_n^{(T^*)}$ ,  $\Lambda_n^{(T^*)}$  und  $W_n^{(T^*)}$  die SVD der verallgemeinerten t-Ergebnisse beschreiben und  $V_n^{(E^*)}$ ,  $\Lambda_n^{(E^*)}$  und  $W_n^{(E^*)}$  die SVD der verallgemeinerten, restlichen Ergebnisse repräsentieren. Die Dimensionen dieser Matrizen sind wie folgt:  $V_n^{(T^*)}$  und  $V_n^{(E^*)}$  sind  $K \times n$ -Matrizen, und  $\Lambda_n^{(T^*)}$ ,  $\Lambda_n^{(E^*)}$ ,  $W_n^{(T^*)}$  und  $W_n^{(E^*)}$  sind  $n \times n$ -Matrizen. Die Spalten der Matrizen  $V_n^{(T^*)}$ ,  $V_n^{(E^*)}$ ,  $W_n^{(T^*)}$  und  $W_n^{(E^*)}$  sind orthogonal, und  $\Lambda_n^{(T^*)}$  und  $\Lambda_n^{(E^*)}$  sind vom Diagonaltyp. Auf der Grundlage der SVD der verallgemeinerten Ergebnismatrizen der Bezugsdaten stellt die folgende Beziehung orthogonale Ergebnisse zur Verfügung:

$$\begin{aligned} \tilde{V}_n^{(T^*)} &= T_n^* W_n^{(T^*)} \Lambda_n^{(T^*)^{-1}} \sqrt{K-1} \\ \tilde{V}_n^{(E^*)} &= E_n^* W_n^{(E^*)} \Lambda_n^{(E^*)^{-1}} \sqrt{K-1} \end{aligned} \quad (13)$$

wobei  $V_n^{(T^*)}$  und  $V_n^{(E^*)}$  orthogonale  $T_n^*$ - und  $E_n^*$ -Ergebnisse mit Einheitsvarianz repräsentieren. Einschließlich Gleichung (13) können die orthogonalen Ergebnisse  $V_n^{(T^*)}$  und  $V_n^{(E^*)}$  direkt aus den vergrößerten Daten und der Fehlermatrix,  $[Y : X]$  und  $[E_n : F_n]$ , folgendermaßen berechnet werden:



$$\begin{aligned}\tilde{v}_n^{(T^*)} &= [Y:X]C_{PLS}^{(n)}W_n^{(T^*)}\Lambda_n^{(T^*)^{-1}}\sqrt{K-1} = [Y:X]G_{PLS}^{(n)} \\ \tilde{v}_n^{(E^*)} &= [E_n:F_n]C_{PLS}^{(n)}W_n^{(E^*)}\Lambda_n^{(E^*)^{-1}}\sqrt{K-1} = [E_n:F_n]H_{PLS}^{(n)}\end{aligned}\quad (14)$$

[0053] Für die verallgemeinerten Ergebnisvektoren des  $i$ -ten Datenpunktes,  $v_n^{(T^*)}$  und  $v_n^{(E^*)}$  kann die Summe der quadrierten Elemente dazu verwendet werden, eine eindimensionale Statistik für jeden Vektor zu definieren. Diese Statistiken werden als  $T^* \cdot T^2$  und  $E^* \cdot T^2$  bezeichnet.  $T^* \cdot T^2$  und  $E^* \cdot T^2$  repräsentieren die EPLS- $T^2$ - und EPLS-Q-Statistik und sind folgendermaßen definiert:

$$\begin{aligned}T^*T_i^2 &= \sum_{j=1}^n y_j \tilde{v}_n^{(T^*)^2} \\ E^*T_i^2 &= \sum_{j=1}^n y_j \tilde{v}_n^{(E^*)^2}\end{aligned}\quad (15)$$

[0054] Unter der Annahme, dass  $v_n^{(T^*)}$  und  $v_n^{(E^*)}$  stochastische Variablen sind, weisen beide Statistiken eine Chi-Quadrat-Verteilung mit  $n$  Freiheitsgraden auf, welche die Vertrauensgrenzen zur Untersuchung dazu zur Verfügung stellt, ob der Prozess sich normal oder anomal verhält. Die Vertrauensgrenzen sind normalerweise so ausgewählt, dass sie 95% und 99% der Population enthalten (EPLS- $T^2$ - oder Q-Werte). Wenn ein neuer EPLS- $T^2$ - oder Q-Wert unterhalb der Grenze liegt, so wird die Hypothese akzeptiert, dass sich der Prozess normal verhält, anderenfalls wird sie zurückgewiesen, und besteht die akzeptierte Hypothese darin, dass sich der Prozess anomal verhält.

[0055] Anomal große EPLS- $T^2$ - und/oder Q-Werte können auftreten, wenn die Beziehung zwischen den Prädiktor- und den Responsvariablen, repräsentiert durch die parametrische Regressionsmatrix, sich geändert hat (beispielsweise ein zeitabhängiger Prozess), oder sich die Störungsstatistiken geändert haben. Andere Gründe können darin bestehen, dass der Prozess an einem unterschiedlichen Arbeitspunkt arbeitet, eine exzessive Variation des Prozesses aufgetreten ist, die nicht in den Bezugsdaten vorhanden war, oder dass ein anomales Prozessverhalten aufgetreten ist. Der Hypothesentest ist daher ein Vergleich zwischen dem momentanen Prozessbetrieb und dem Prozessbetrieb, der durch die Bezugsdaten wiedergegeben wird. Es wird darauf hingewiesen, dass die Bezugsdaten den Prozess im Normalbetrieb beschreiben, und jede Variation wiedergeben müssen, die im Normalbetrieb auftreten kann, da anderenfalls der statistische Hypothesentest zu empfindlich wird.

[0056] [Wise, 1996] betonte, dass insbesondere die  $T^2$ -Statistik nicht normal verteilt sein kann. [Dunia, 1996] untersuchte den Einfluss eines exponentiell gewichteten, beweglichen Mittelwerts (EWMA) auf die Q-Statistik, welche PCA verwendet. Es stellte sich heraus, dass die mittlere Laufzeit (ARL), nämlich die mittlere Zeit, die verstreicht, bis ein anomales Prozessverhalten festgestellt wird, zur Feststellung von Fehlerzuständen bei Sensoren dadurch verringert werden kann, dass die EWMA-Q-Statistik verwendet wurde. In der vorliegenden Beschreibung wird eine EWMA-Vorgehensweise bei der PLS-Q- und der EPLS-Q-Statistik angewendet. Aus diesen Gründen wird jede Vertrauensgrenze empirisch in der vorliegenden Beschreibung bestimmt, wie dies bereits vorgeschlagen wurde [Box, 1978].

[0057] Die Diagnose des festgestellten, anomalen Prozessverhaltens kann durch Untersuchung der Residuen der Responsvariablen durchgeführt werden. Diese Residuen können zu jedem Zeitpunkt in einem Balkendiagramm aufgetragen werden. Von einem großen Residuum einer bestimmten Responsvariablen wird angenommen, dass es durch das anomale Prozessverhalten beeinflusst wird, und umgekehrt. Weiterhin führte [Kourti, 1995] aus, dass ein Balkendiagramm auch durch die Residuen der Prädiktorvariablen zu jedem Zeitpunkt erzeugt werden kann. Wenn das Residuum einer bestimmten Prädiktorvariablen größer ist, dann wird auch bei dieser Variablen angenommen, dass sie ebenfalls durch das anomale Prozessverhalten beeinflusst wird. Die "Größe" der Residuen einer bestimmten Prädiktor- oder Responsvariablen ist relativ in Bezug auf die Residuen anderer Prädiktor- und Responsvariablen. Ein Vergleich der momentanen Residuen muss ebenfalls relativ zu den Residuen der Bezugsdaten durchgeführt werden. Die Balkendiagramme werden auch als Fehlerbeitragsdiagramme (EC-Diagramme) bezeichnet, eines für die Responsvariablen und eines für die Prädiktorvariablen. Die Balkenhöhen repräsentieren daher die quadrierten Residuen der Responsvariablen und die quadrierten Residuen der Prädiktorvariablen. Um diese Werte miteinander statistisch vergleichen zu können, muss eine Normierung durchgeführt werden. Erfolgt dies nicht, verursacht eine Responsvariable, die nicht so gut vorhergesagt werden kann wie andere, beispielsweise im Mittel größere Balken in Bezug auf die anderen Responsvariablen, und umgekehrt.

#### 4. FALLUNTERSUCHUNGEN

##### 4.1 Katalytische Fluidkrackeinheit

[0058] Eine katalytische Fluidkrackeinheit oder FCCU ist eine ökonomisch bedeutsame Einheit bei Raffiniervorgängen. Sie empfängt typischerweise mehrere verschiedene, schwere Rohstoffe von anderen Raffiniereinheiten und krackt diese Ströme zur Erzeugung leichterer, wertvollerer Komponenten, die schließlich zu Benzin und anderen Produkten gemischt werden. Die spezielle Einheit Modell IV, die von [McFarlane, 1993] beschrieben wurde, ist in Fig. 1 dargestellt. Der hauptsächliche Rohstoff der Einheit ist Gasöl, jedoch tragen auch Ströme aus schwererem Diesel und Waschöl zum gesamten Rohstoffstrom bei. Frischer Rohstoff wird in einem Wärmetauscher und einem Ofen vorerwärmt und dann dem Riser zugeführt, indem er mit heißem, regeneriertem Katalysator von dem Regenerierer gemischt wird. Eine Aufschlammung von dem Hauptfraktionierkolonnenboden wird ebenfalls zum Riser recycelt. Der heiße Katalysator stellt die Wärme zur Verfügung, die für die endothermen Crackreaktionen benötigt wird. Die gasförmigen, gekrackten Pro-

dukte werden an die Hauptfraktionierkolonne zur Trennung geleitet. Feuchtgas von der Oberseite der Hauptfraktionierkolonne wird auf den Druck der Light-Ends-Anlage durch den Feuchtgaskompressor erhöht. Eine weitere Trennung leichter Komponenten tritt in diesem Light-Ends-Trennabschnitt auf.

[0059] Die ausgewählten Prädiktorvariablen für die FCCU-Fallstudie sind in Tabelle 1 angegeben.

Tabelle 1

Ausgewählte Prädiktorvariablen für die FCCU-Fallstudie

PRÄDIKTORVARIABLEN	SIGNAL
Waschölflußrate	zeitlich konstant
Diesel-Flussrate	ARMA-Sequenz
gesamter frischer Rohstoff	ARIMA-Sequenz
Aufschlammungs-Flussrate	ARIMA-Sequenz
Vorheizer-Auslasstemperatur	ARIMA-Sequenz
Reaktoreinstellpunkt	konstant
Feuchtgaskompressor-Saugventilposition	abhängig vom Reaktordruck

[0060] Alle diese Variablen gehören zum Rohstoffabschnitt der Einheit. Um realistische Störbedingungen zu simulieren, wurden verschiedene Signale des Typs Autoregressive Integrated Moving Average (ARIMA) diesen Variablen überlagert, mit der Ausnahme der Dieselflussrate, die nur ein Signal des Typs Autoregressive Moving Average (ARMA) empfing, und mit Ausnahme der Waschölflußrate und des Reaktoreinstellpunkts, die immer konstant oder gleich Null waren.

[0061] Die Responsgruppe umfasste überschüssigen Sauerstoff in dem Rauchgas, Konzentration von Kohlenmonoxid in dem Rauchgas, Riser-Temperatur, Regeneratorbettemperatur, Regeneratorsteigrohrpegel sowie neun weitere gemessene Variablen des Systems, vergleiche [McFarlane, 1993] bezüglich einer vollständigen Liste gemessener Variablen für das FCCU-System.

[0062] Um die PLS- und EPLS-Algorithmen zu testen, wurde der FCCU-Simulator vergrößert, so dass er verschiedene vorprogrammierte Fehler enthielt, die auf Befehl eingesetzt werden konnten. Der erste war eine stufenweise Änderung auf den Koksablagerungsfaktor des Rohstoffs, wodurch ein Pfropf aus einem schwerer als normalen Rohstoff simuliert wurde, der in die Einheit hineingelangt. Der zweite simulierte eine Unterbrechung des Flusses des regenerierten Katalysators zwischen dem Regenerierer und dem Riser, was typischerweise durch teilweise oder vollständige Verstopfung von Dampfinjektoren hervorgerufen wird, die sich in dieser Leitung befinden.

[0063] Beim ersten Durchlauf war kein fortgeschrittenes Regelsystem vorhanden, nur Regelsteuerungen für den Reaktor, Luftkompressorflussraten und den Reaktordruck. Bei dieser Regelkonfiguration war keine Rückkopplung zwischen Respons- und Prädiktorvariablen bei den Läufen 1 und 2 vorhanden, da die Feuchtgaskompressor-Saugventilposition weggelassen wurde.

[0064] Fig. 3 zeigt die PLS- $T^2$ - und Q-Überwachungsdiagramme sowie die entsprechenden EPLS-Diagramme über einen Zeitraum von annähernd 1500 Stunden im Normalbetrieb. In allen Figuren, in denen  $T^2$ - und Q-Plots gezeigt sind, gibt die obere durchgezogene Linie die Vertrauensgrenze von 99% für die jeweilige aufgetragene Statistik an, während die untere gestrichelte Linie die Vertrauensgrenze von 95% angibt. Weiterhin ist die Ordinate jedes  $T^2$ - und Q-Plots im Zehnerlogarithmus dargestellt. Die Abtastperiode wurde auf 30 Minuten eingestellt.

[0065] In Fig. 4 sind die Responses der  $T^2$ - und Q-Statistiken von PLS und EPLS für den ersten Fehler gezeigt, der bei annähernd 190,5 Stunden auftritt. Da dieser Fehler eine Änderung der Zusammensetzung des Rohstoffes simuliert, nämlich einen Pfropf aus schwererem Rohstoff, wirkt er sich sofort in dem Riser aus, und danach in anderen Teilen der Einheit, die durch eine Änderung des Riser-Zustands beeinflusst werden, nach 191 Stunden. Allerdings gibt es keinen direkten, mechanischen Weg zurück zu einem Teil des Rohstoffsystems, und daher wird keine der Prädiktorvariablen, wie sie für die Läufe 1 und 2 definiert wurden, direkt beeinflusst. Ebenso wenig werden sie durch Rückkopplung von den Responsvariablen beeinflusst, da kein fortgeschrittenes Regelsystem vorhanden ist, um eine derartige Rückkopplung zur Verfügung zu stellen. Daher ergibt sich aus den PLS- $T^2$ - und Q-Diagrammen überhaupt keine Anzeige, dass der Fehler einen anomalen Zustand hervorgerufen hat.

[0066] Im Gegensatz hierzu identifizieren die EPLS- $T^2$ - und Q-Statistiken, die in Fig. 4 aufgetragen wurden, deutlich den anomalen Zustand an der Vertrauensgrenze von 99%. Das EC-Diagramm, entsprechend der Zeit, zu welcher dieses Ereignis deutlich wird (nach 191 Stunden), ist in Fig. 5 gezeigt. Die Variablen 12 und 13, nämlich überschüssiger Sauerstoff in dem Rauchgas bzw. Konzentration von Kohlenmonoxid in dem Rauchgas, tragen deutlich zu dem Ereignis bei. Dies macht physikalisch Sinn, da ein Pfropf aus einem schwereren Rohstoff zu einem rapiden Anstieg der Menge an Koks führt, die auf dem Katalysator in dem Riser abgelagert wird, und zum Regenerierer transportiert wird, mit einer direkten Auswirkung auf den Sauerstoffverbrauch und die Erzeugung von Kohlenmonoxid. Das Beitragsdiagramm ergibt nicht direkt die potentielle Quelle des Fehlers, sondern stellt einem erfahrenen Anlagenbetreiber Information zur Verfügung, die dazu hilft, potentielle Ursachen einzuzugrenzen. Im Gegensatz zu dem EC-Diagramm für die Responsvariablen

zeigt das EC-Diagramm für die Prädiktorvariablen keinen großen Beitrag für irgendeine Variable.

- [0067] Bei dem zweiten Lauf wurde der Fehler des regenerierten Katalysators nach 322 Stunden eingesetzt. Da die Prädiktorvariablen alle aus dem Rohstoffabschnitt der Einheit stammen, ergibt wiederum ein Fehler oder eine Störung, die entweder im Reaktor oder dem Regenerierer auftreten, oder in den Verbindungskatalysatorleitungen, keinen mechanistischen Weg zurück zu diesen Variablen. In diesem Fall beeinflusst der Fehler nur Responsvariablen und stellen herkömmliche PLS-T<sup>2</sup>- und Q-Diagramme das Ereignis nicht fest. Dies wird aus den oberen beiden Plots von Fig. 6 deutlich. Die EPLS-T<sup>2</sup>- und Q-Diagramme stellen jedoch den anomalen Zustand deutlich nach 322,5 Stunden fest. Das entsprechende EC-Diagramm für die Responsvariablen, das in Fig. 7 gezeigt ist, zeigt an, dass überschüssiger Sauerstoff in dem Rauchgas und der Standrohrpegel in dem Regenerierer signifikante Beiträge liefern. Dies lässt sich einfach erklären, da jede Änderung des Flusses des regenerierten Katalysators das Materialgleichgewicht in dem Standrohr beeinflusst, und daher dessen Pegel. Eine Änderung des Flusses des regenerierten Katalysators beeinflusst auch das Verhältnis von Katalysator zu Rohstoff in dem Riser, was zu einer Änderung der Menge an Koks führt, die auf verbrauchtem Katalysator abgelagert wird, und dann zu einer Änderung des Pegels des Sauerstoffverbrauchs in dem Regenerierer. Es wird darauf hingewiesen, dass das EC-Diagramm für die Prädiktorvariablen keinen anomal großen Beitrag für irgendeine Variable zeigt, da keine Regelungsrückkopplung vorhanden ist.
- [0068] Für die Läufe 3 und 4 wurde die Position des Feuchtgaskompressorausgvents als Prädiktorvariable mit berücksichtigt. Die Auswirkung irgendeiner Störung oder irgendeines Fehlers, welche den Reaktordruck beeinflussen, wird daher an die Prädiktorvariable übertragen, die durch die Rückkopplungswirkung der Reaktordruckregelung festgelegt wird. In diesem Fall würde man sowohl bei PLS als auch bei EPLS erwarten, dass sie einen anomalen Zustand feststellen, der sich bei dieser Art von Fehler ergibt, und dies zeigt sich deutlich in Fig. 8 für den ersten Fehler, und in Fig. 9 für den zweiten Fehler.
- [0069] Der Einsatz der Vorgehensweise I bei der FCCU-Falluntersuchung hat gezeigt, dass die PLS-Überwachungsdiagramme nur ein anomales Prozessverhalten feststellen können, wenn eine Steuerungsrückkopplung bei den Prädiktorvariablen vorhanden ist. Dies führt weiter dazu, dass die EC-Diagramme für die Prädiktorvariablen keinen anomal großen Beitrag zu irgendeiner Variablen zeigen. Im Gegensatz hierzu stellten die EPLS-Diagramme beide Fehler fest. Beim Vorhandensein einer Steuerungsrückkopplung sind jedoch auch die PLS-Diagramme empfindlich.

#### 4.2 Fließbettreaktor

- [0070] Dieser industrielle Prozess erzeugt zwei Lösungsmittelchemikalien, die mit F und G bezeichnet sind, und besteht aus mehreren Betriebseinheiten. Die Hauptelemente dieser Anlage sind kontinuierlich arbeitende Einheiten, in denen die chemische Reaktion durchgeführt wird. Diese Einheiten sind fünf parallel arbeitende Fließbettreaktoren, in denen jeweils F und G durch komplizierte, exothermische, chemische Reaktionen erzeugt werden. Diese Reaktoren werden mit zwei unterschiedlichen Strömen aus fünf verschiedenen Reaktanten versorgt. Fig. 2 zeigt schematisch einen Reaktor und seine benachbarten Einheiten.
- [0071] Der erste Strom besteht aus dem Reaktanden A, und der zweite Strom aus den Reaktanden B, C, D und E. A und B sind Moleküle des Typs X<sub>2</sub>, C ist eine Säure, D sind Moleküle, die in stromaufwärtigen Einheiten erzeugt werden, und E sind recycelte Erzeugnisse der Anlage. D und E werden durch einen stromaufwärtigen Verdampfer verdampft, bevor sie in den Reaktor als Teil des zweiten Stroms hineingelangen.
- [0072] Schließlich wird nach Verlassen des Reaktors die Trennung von F und G durch stromabwärtige Destilliereinheiten erreicht.
- [0073] Der Reaktor besteht aus einem großen Mantel und einer Anzahl vertikal angeordneter Rohre, in denen die chemische Reaktion durchgeführt wird, unterstützt durch einen Fließbettkatalysator. Ein Thermoelement ist am Boden jedes Rohrs angebracht, um die Temperatur des Fließbettes zu messen. Zum Abführen der Wärme der exothermen Reaktion wird Öl um die Rohre herum umgewälzt. Das Verhältnis von F : G wird regelmäßig in einem Labor analysiert. Auf der Grundlage dieser Analyse wird das Verhältnis von F : G durch die Reaktorrohstoffzufuhrraten eingestellt. Damit der Katalysator ständig im Fließbett vorhanden ist, wird die Fließbettesgeschwindigkeit dadurch konstant gehalten, dass der Reaktordruck in Bezug auf die Gesamtflussrate eingestellt wird.
- [0074] Die chemische Reaktion wird durch nicht gemessene Störungen und Änderungen des Fließbettes des Katalysators beeinflusst. Die am häufigsten beobachtete, nicht gemessene Störung wird durch Umschlagen des Drucks des Dampfstroms hervorgerufen, der vom Verdampfer benötigt wird. Ungemessene Störungen können auch durch das Kühlmittel (Öl) hervorgerufen werden, das von einer getrennten Einheit zur Verfügung gestellt wird. Infolge des Regelschemas der Verdampfereinheit führen die Druckumschläge des Dampfstroms zu einer größeren oder kleineren Flussrate des zweiten Stroms, der in den Reaktor eintritt. Fließbettprobleme treten auf, wenn die Katalysatorverteilung über ein Rohr am Boden des Rohrs erheblich größer ist. Dies bringt es mit sich, dass die chemische Reaktion am Boden des Rohrs verstärkt wird, was zu einer signifikanten Erhöhung der Rohrtemperatur führt.
- [0075] Während eines Zeitraums von mehreren Wochen wurden normale Betriebsdaten sowie Daten, welche Prozessanomalitäten enthielten, für einen bestimmten Reaktor erhalten. Die Datengruppe zum Aufnehmen des normalen Prozessbetriebs (Bezugsdaten) musste sorgfältig ausgewählt werden. Es musste sichergestellt werden, dass die Bezugsdaten keine Störungen einfangen, wie sie voranstehend beschrieben wurden, oder Fließbettprobleme eines oder mehrerer Rohre. Weiterhin könnte, wenn die Menge an Bezugsdaten zu gering wäre, eine normale Variation, die während der chemischen Reaktion auftritt, nicht in der Beschreibung enthalten sein. Jede Datengruppe beschreibt den Prozess im Dauerzustandsbetrieb.
- [0076] Zum Identifizieren eines Dauerzustands-PLS-Modells mussten Prädiktor- und Responsvariablen ausgewählt werden. Die Prädiktorvariablen sind die Flussrate der Reaktanden A, B, D und E, der Dampfstrom zum Verdampfer, und ein zusätzlicher Strom, der zur Verringerung des Drucks in dem Verdampfer erforderlich ist. Als Responsvariable wurde die Temperatur jedes Rohrs ausgewählt.
- [0077] Eine Voruntersuchung der Daten ergab, dass die Rohrtemperaturen stark korreliert sind. Weiterhin ist eine Kor-

relation auch zwischen den Prädiktorvariablen vorhanden. Die Bestimmung der Anzahl an LVs, die beibehalten werden sollen, ergab jedoch, dass sämtliche sechs LVs signifikant zur Vorhersage der Responsmatrix beitragen. Die Auswahl, wieviele LVs beibehalten werden sollen, wurde dadurch durchgeführt, dass eine Kreuzüberprüfung so durchgeführt wurde, dass jeweils eine Größe weggelassen wurde. Dieser Fall stellt daher ein Beispiel für den zweiten Nachteil dar, der auftritt, wenn die Prozesszustandsüberwachung mit der Vorgehensweise I durchgeführt wird.

[0078] Obwohl jeder Ergebnisvektor der Prädiktormatrix beibehalten werden muss, verringert PLS die Anzahl an Prozessvariablen beträchtlich. Nachdem die Identifizierungsprozedur beendet wurde, wurden die sechs verallgemeinerten t-Ergebnisse und die verallgemeinerten Residuenergebnisse für die Bezugsdaten gemäß Gleichung (9) berechnet. Hieran schloss sich die Berechnung der entsprechenden Werte der PLS- $T^2$ -, EPLS- $T^2$ - und Q-Statistik und der Schwellenwerte für die zugehörigen Überwachungsdiagramme an. Es wird darauf hingewiesen, dass die PLS-Q-Statistik nicht bestimmt werden kann.

[0079] Fig. 12 zeigt die Überwachungsdiagramme für die PLS- und die EPLS-Vorgehensweise für die Bezugsdaten. Es wird darauf hingewiesen, dass die Werte jeder Statistik im logarithmischen Maßstab dargestellt sind. Die Diagramme der PLS- $T^2$ - und der EPLS- $T^2$ -Statistik zeigen eine natürliche Variation des Prozesses, beispielsweise infolge von Änderungen der Zufuhr. Weiterhin zeigt das Diagramm der EPLS- $T^2$ -Statistik den Einfluss gemeinsamer Variationen (nicht gemessener Störungen), die das Modell nicht beschreiben kann.

[0080] Das erste beobachtete, anomale Prozessverhalten repräsentiert eine große, nicht gemessene Störung infolge des Abfalls des Dampfdrucks. Die sich ergebenden Überwachungsdiagramme sind in Fig. 13 dargestellt. Obwohl die Dampfdruck konstant bleibt, ändert sich das Enthalpiegleichgewicht innerhalb des Verdampfers und beeinflusst die Zusammensetzung der Reaktanten D und E in dem zweiten Strom zum Reaktor. Die nicht gemessene Störung trat nach etwa 1300 Minuten in der aufgezeichneten Datengruppe auf. Die EPLS-Q-Statistik stellt diese nicht gemessene Störung unmittelbar danach fest, da die Zusammensetzung des zweiten Stroms deutlich die Reaktionsbedingungen beeinflusst, welche das PLS-Modell nicht beschreiben kann. Die nicht gemessene Störung wird nicht von der PLS- $T^2$ - und der EPLS- $T^2$ -Statistik aufgenommen, da die Prozessvariation dieses Ereignisses nicht die Variationen in den Bezugsdaten überschreitet. Die entsprechenden EC-Diagramme für 1500, 1501 und 1502 Minuten, vergleiche Fig. 4, zeigen, dass bei etwa der Hälfte der Rohre die Temperatur nicht exakt in Bezug auf die Bezugsdaten vorhergesagt werden kann. Die nicht gemessene Störung beeinflusst deutlich die Reaktionsbedingungen in sämtlichen Rohren, was dadurch bestätigt werden konnte, dass hintereinander die drei EC-Diagramme untersucht wurden. Die Diagnose dieses anomalen Prozessverhaltens ergibt sich für einen erfahrenen Anlagenbetreiber, der aus der bereitgestellten Information zurück auf den Abfall des Dampfdrucks schließen kann.

[0081] Das zweite anomale Prozessverhalten beschreibt ein Fließbettproblem in einem der Rohre. Es gibt einige Manipulationen, die ein Anlagenbetreiber durchführen kann, um das Fließbett zu verbessern und daher die Rohrtemperatur zurück auf ihren normalen Betriebswert zu bringen. Die erste Temperaturerhöhung wurde jedoch von dem Anlagenbetreiber nicht bemerkt. Als die zweite Temperaturerhöhung festgestellt wurde, wurde der Versuch unternommen, die Temperatur zurück zum Normalbetrieb zu bringen. Fig. 15 zeigt die entsprechenden PLS- $T^2$ - und EPLS-Überwachungsdiagramme. Beide EPLS-Statistiken stellen in beiden Fällen fest, dass die Rohrtemperatur anomal hoch ist. Im Gegensatz hierzu löst die PLS- $T^2$ -Statistik nur einen Alarm an der Vertrauensgrenze von 99% nach dem zweiten Temperaturanstieg aus. Auch die PLS- $T^2$ -Statistik überschreitet zumindest die Vertrauensgrenze von 95%, und zeigt daher ein anomales Prozessverhalten an. Die Empfindlichkeit des PLS- $T^2$ -Diagramms ist deswegen vorhanden, da die Rückkopplung der Regelschleife für die Fließbettgeschwindigkeit infolge des anomalen Rohrverhaltens vorhanden ist. Die Manipulation des Prozessbetreibers lässt sich aus dem deutlichen Kurvenknick in jedem Überwachungsdiagramm etwa bei 900 Minuten in Fig. 15 erkennen. Nachdem dieser Versuch fehlschlug, schaltete der Anlagenbetreiber schließlich das Rohr ab. Die Überwachungsdiagramme der EPLS- $T^2$ - und der Q-Statistik entsprechen dem Abschalten durch Stilllegen. Fig. 16 zeigt die EC-Diagramme nach 436 Minuten, 880 Minuten, 905 Minuten und 910 Minuten. Entsprechend den EC-Diagrammen könnte der Anlagenbetreiber mehr als 450 Minuten früher Maßnahmen eingeleitet haben, um den Betrieb dieses Rohrs aufrechtzuerhalten.

[0082] Die Anwendung der Vorgehensweise I bei diesem industriellen Beispiel zeigte, dass PLS-Überwachungsdiagramme unempfindlich (zumindest beim ersten Beispiel) in der Hinsicht sind, ein anomales Verhalten des Prozesses festzustellen. Bei dem zweiten Beispiel konnte jedoch PLS das anomale Rohrverhalten feststellen (zumindest bei der Vertrauensgrenze von 95%). Im Gegensatz hierzu konnten die EPLS-Überwachungsdiagramme deutlich jede Prozessanomalie feststellen (an der Vertrauensgrenze von 99%). Weiterhin musste jeder Wert für LV beibehalten werden, infolge seiner Signifikanz zum Vorhersagen der Responsvariablen. Daher konnte nur ein Überwachungsdiagramm bei der Vorgehensweise I erhalten werden.

[0083] Die zweite Vorgehensweise (Vorgehensweise II) würde zu 16 Streuplots führen und sechs Überwachungsdiagrammen, von denen keines eine Variation der Responsvariablen beschreibt. Im Gegensatz hierzu benötigt EPLS nur zwei Überwachungsdiagramme und beschreibt die EPLS- $T^2$ -Statistik eine Variation der Responsvariablen, vergleiche Gleichung 9.

## 5. SCHLUSSFOLGERUNGEN

[0084] Bei der vorliegenden Beschreibung werden die herkömmlichen PLS-Vorgehensweisen für die Zustandsüberwachung kontinuierlicher, industrieller Prozesse, wie sie in [MacGregor, 1991, 1995], [Kresta, 1991] und [Wise, 1996] beschrieben werden, erneut überprüft, und werden Problembereiche hervorgehoben. Aus dieser Untersuchung ergibt sich, dass die herkömmlichen PLS-Überwachungsdiagramme entweder unempfindlich oder schwierig zu analysieren sein können, falls sich der Prozess anomal verhält. Diese Beschreibung stellt eine Ausweitung des Standard-PLS-Algorithmus zur Verfügung, bezeichnet als EPLS, die zu der Definition neuer PLS-Ergebnisse führt, die als verallgemeinerte Ergebnisse bezeichnet werden. Ähnlich wie bei herkömmlichen PLS-Vorgehensweisen, können Statistiken auf der Grundlage der verallgemeinerten Ergebnisse von EPLS definiert werden, die in Abhängigkeit von der Zeit auf Überwachungs-

Diagrammen aufgetragen werden können. Diese Überwachungsdiagramme beschreiben die Gesamtvariation der Prädiktor- und Responsvariablen (EPLS-T<sup>2</sup>-Diagramm) und ihrer Residuen (EPLS-Q-Diagramm).

[0085] Eine theoretische Analyse der Überwachungsdiagramme, die aus den verallgemeinerten Ergebnissen von EPLS abgeleitet werden, und herkömmlicher PLS-Vorgehensweisen ergibt Folgendes:

1. Bei der Vorgehensweise I wird, wenn ein anomales Verhalten Responsvariablen beeinflusst, die nicht Teil einer geschlossenen Regelschleife sind, diese Situation möglicherweise nicht festgestellt. Bei EPLS wird die anomale Variation der Responsvariablen deutlich.
2. Falls die Responsvariablen stark korreliert sind, jedoch die Prädiktorvariablen nicht, erzeugt die Vorgehensweise I nur ein Diagramm, das T<sup>2</sup>-Diagramm. Bei EPLS bleiben beide Diagramme relevant, unabhängig von der Anzahl an latenten Variablen, die beibehalten werden.
3. Die zweite Vorgehensweise (Vorgehensweise II) definiert eine Anzahl an Diagrammen in Abhängigkeit von der Anzahl an latenten Variablen, die beibehalten werden. Im Gegensatz hierzu benötigt EPLS nur zwei Diagramme, unabhängig von der Auswahl ausgewählter, latenter Variablen.
4. Der Einsatz des Vorhersagefehlerquadratdiagramms zusammen mit den beiden Überwachungsdiagrammen der Vorgehensweise I kann die voranstehend geschilderten Schwierigkeiten überwinden. Allerdings führt dies zu zumindest drei Überwachungsdiagrammen, und ist die Variation der Responsvariablen in keinem dieser Diagramme vorhanden. Im Gegensatz hierzu benötigt EPLS nur zwei Überwachungsdiagramme, und wird die Variation der Responsvariablen nicht in dem EPLS-T<sup>2</sup>-Diagramm akkumuliert.

[0086] Die vorliegende Beschreibung stellt auch zwei Einsatzuntersuchungen zur Verfügung, um die voranstehenden, theoretisch abgeleiteten Ergebnisse zu bestätigen. Diese Anwendungen betreffen die Simulation einer katalytischen Fluidkrackeinheit (FCCU) und einen realen industriellen Prozess. Zwei anomale Situationen sind in beiden Falluntersuchungen vorhanden, welche die Auswirkung einer nicht gemessenen Störung und eine "interne" Änderung des Prozessverhaltens beschreiben.

[0087] Die Ergebnisse der FCCU-Falluntersuchung zeigen deutlich, dass eine Regelungsrückkopplung für die Vorgehensweise I unbedingt erforderlich ist, um ein robustes und empfindliches Zustandsüberwachungswerkzeug zur Verfügung zu stellen. Falls dies nicht garantiert ist, kann es bei dieser Vorgehensweise auftreten, dass anomale Situationen nicht festgestellt werden. Im Gegensatz hierzu stellt die EPLS-Vorgehensweise robuste und empfindliche Überwachungsdiagramme zur Verfügung, unabhängig von dem Vorhandensein einer Regelungsrückkopplung.

[0088] Dies kann auch durch die Untersuchung des industriellen Falls bestätigt werden. Die erste anomale Situation wird nicht von der PLS-T<sup>2</sup>-Statistik festgestellt, da der Umschlag des Dampfdrucks nicht die Prädiktorvariablen stark genug beeinflusst, um festgestellt zu werden. Es wird allerdings der Zustand der chemischen Reaktion innerhalb der Rohre beeinflusst und daher das Verhalten der Responsvariablen. Die zweite Situation beschreibt ein anomales Verhalten eines der Rohre. In dieser Situation beeinflusst die Regelungsrückkopplung die Prädiktorvariablen infolge des anomalen Verhaltens.

[0089] Die Untersuchung des industriellen Falls zeigt weiterhin, dass sämtliche latenten Variablen signifikant zur Vorhersage der Responsvariablen beitragen. Daher ist nur ein Überwachungsdiagramm (das T<sup>2</sup>-Diagramm) für die Vorgehensweise I verfügbar. Weiterhin würde die Vorgehensweise II zu insgesamt 16 Streuplots und 6 Überwachungsdiagrammen führen. Bei dieser Anzahl an relevanten Diagrammen ist es entsprechend mühsam, ein anomales Verhalten festzustellen, im Unterschied zur Beobachtung der beiden EPLS-Überwachungsdiagramme.

[0090] Die weitere Forschung in Bezug auf die verallgemeinerten Ergebnisse von EPLS konzentriert sich auf Anwendungen, welche dynamische Prozessmodelle verwenden, wie sie beispielsweise für eine Modellvorhersageregulation erforderlich sind. Es soll untersucht werden, ob dieses dynamische Prozessmodell auch die Grundlage für die Prozesszustandsüberwachung bereitstellt, im Unterschied zu der Dauerzustandsuntersuchung, die bislang vorgenommen wurde. Darüber hinaus wird der Einsatz der verallgemeinerten EPLS-Ergebnisse für die Überwachung von postenweise ablaufenden Prozessen der Gegenstand zukünftiger Überlegungen sein. Die PLS-Vorgehensweise zur Überwachung postenweise ablaufender Prozesse, die in [Nomikos, 1994 und 1995] diskutiert wurde, wird daher die Grundlage für die Unterscheidung eines "guten" Postens von einem "schlechten" Posten sein.

## ANHÄNGE

### A.1 Theoretische Analyse des PLS-Algorithmus

[0091] Der PLS-Identifizierungsalgorithmus beruht auf der Bestimmung jedes Paares von Komponentenmatrizen  $X_k$  und  $Y_k$  (siehe Gleichung 1), durch eine iterative Prozedur. Nachdem der k-te Iterationsschritt durchgeführt wurde, werden die berechneten Komponentenmatrizen von der Prädiktormatrix bzw. der Responsmatrix subtrahiert, vor der Berechnung des (k+1)-ten Iterationsschrittes. Die Subtraktion der Komponentenmatrizen wird auch als Druckablassprozedur bezeichnet und verläuft folgendermaßen:

$$\begin{aligned} X_k &= X_{k-1} - \tilde{X}_k = X_{k-1} - t_k p_k^T \\ Y_k &= Y_{k-1} - \tilde{Y}_k = Y_{k-1} - \hat{u}_k q_k^T \end{aligned} \quad (A1)$$

[0092] Die Ergebnisvektoren  $t_k$ ,  $\hat{u}_k$  und Belastungsvektoren  $p_k$  und  $q_k$  werden dazu bestimmt, um den Beitrag jedes Paares von Komponentenmatrizen zu den Prädiktor- und Responsmatrizen zu maximieren. Dies wird dadurch erreicht, dass folgende Kriterien erfüllt werden:

$$\begin{aligned}
 t_i &= X_{i-1} w_i; & \|w_i\|_2^2 - 1 &= 0 \\
 u_i &= Y_{i-1} v_i; & \|v_i\|_2^2 - 1 &= 0, \\
 J_w &= \max\{t_i^T u_i\} = \max\{w_i^T \cdot X_{i-1}^T \cdot Y_{i-1} v_i\}
 \end{aligned} \tag{A2}$$

$$\begin{aligned}
 u_i &= b_i t_i + e_i = \hat{u}_i + e_i \\
 J_e &= \min\{e_i^T e_i\} = \min\{u_i^T u_i - 2b_i t_i^T u_i + b_i^2 t_i^T t_i\} \text{ und}
 \end{aligned} \tag{A3}$$

$$\begin{aligned}
 J_p &= \min\{trace\{[X_{i-1} - t_i p_i^T]^T [X_{i-1} - t_i p_i^T]\}\} \\
 J_q &= \min\{trace\{[Y_{i-1} - \hat{u}_i q_i^T]^T [Y_{i-1} - \hat{u}_i q_i^T]\}\}
 \end{aligned} \tag{A4}$$

[0093] Lösungen für die drei Kostenfunktionen müssen daraufhin berechnet werden. Beginnend mit Gleichung (A2) werden  $w_k$  und  $v_k$  als Gewichtsvektor der Prädiktormatrix bzw. der Responsmatrix bezeichnet, und repräsentiert  $J_{wv}$  den Wert der entsprechenden Kostenfunktion. Nach [Hoskuldsson, 1988] kann einschließlich der Randbedingungen für die Länge der Gewichtsvektoren Gleichung (A2) folgendermaßen umgeschrieben werden:

$$\begin{aligned}
 X_{i-1}^T Y_{i-1} Y_{i-1}^T X_{i-1} w_i &= \lambda_i w_i \\
 Y_{i-1}^T X_{i-1} X_{i-1}^T Y_{i-1} v_i &= \lambda_i v_i
 \end{aligned} \tag{A5}$$

[0094] Aus Gleichung (A5) wird deutlich, dass  $w_k$  der Eigenvektor ist, der dem größten Eigenwert der Kreuzcovarianzmatrix

$$X_{k-1}^T \cdot Y_{k-1} Y_{k-1}^T \cdot X_{k-1} \tag{A6}$$

zugeordnet ist und  $v_k$  der Eigenvektor ist, der dem größten Eigenwert der Kreuzcovarianzmatrix

$$Y_{k-1}^T \cdot X_{k-1} X_{k-1}^T \cdot Y_{k-1} \tag{A7}$$

$$X_{k-1}^T \cdot Y_{k-1} Y_{k-1}^T \cdot X_{k-1} \tag{A8}$$

entspricht. In Gleichung (A3) ist  $b_k$  der Regressionskoeffizient zwischen dem  $k$ -ten Paar von Ergebnisvektoren, nämlich  $t_k$  und  $u_k$ , und ist  $J_q$  der Wert der Kostenfunktion. Die Lösung von Gleichung (A3) ist die übliche Lösung der kleinsten Fehlerquadrate für  $b_k$ . Für Gleichung (A4) ist  $J_p$  die zugehörige Kostenfunktion zur Bestimmung von  $p_k$ , und von  $J_q$  zur Berechnung von  $q_k$ . Die Kostenfunktionen  $J_p$  und  $J_q$  werden durch die folgende Lösung minimiert [Geladi, 1986]

$$\begin{aligned}
 p_i &= \frac{X_{i-1}^T t_i}{t_i^T t_i} \\
 q_i &= \frac{Y_{i-1}^T \hat{u}_i}{\hat{u}_i^T \hat{u}_i}
 \end{aligned} \tag{A6}$$

[0095] Schließlich kann unter Einfluss von  $n$  latenten Variablen die Matrix der Regressionskoeffizienten

$$B_{PLS}^{(n)} \tag{A7}$$

zwischen der Prädiktormatrix und der Responsmatrix folgendermaßen [Lindgren, 1993] berechnet werden:

$$\begin{aligned}
 Y_0 &= X_0 \cdot B_{PLS}^{(n)} + E_n \\
 B_{PLS}^{(n)} &= W_n [P_n^T W_n]^{-1} diag\{b_n\} Q_n^T
 \end{aligned} \tag{A7}$$

wobei  $W_n$ ,  $P_n$  und  $Q_n$  Matrizen sind, in denen die  $n$  Gewichtsvektoren  $w_k$  und Belastungsvektoren  $p_k$  und  $q_k$  als Spalten gespeichert sind. Gemäß Gleichung (A2) wird der Gewichtsvektor für die Prädiktormatrix,  $w_k$ , mit der Druckablass-Prädiktormatrix  $X_{k-j}$  multipliziert, um den Ergebnisvektor  $t_k$  zu bestimmen. Allerdings gab [Lindgren, 1993] an, dass der Ergebnisvektor  $t_k$  auch direkt aus der ursprünglichen Prädiktormatrix  $X_0$  folgendermaßen berechnet werden kann:

$$t_k = X_0 \cdot r_k; \quad r_k = \left[ \prod_{i=1}^{k-1} [\mathcal{I} - w_i \cdot p_i^T] \right] \cdot w_k \quad (A8)$$

$$T_n = X_0 [r_1 \quad \dots \quad r_n] = X_0 R_n; \quad R_n = W_n [P_n^T W_n]^{-1}$$

#### A.2 Für Gleichung (7) erforderlicher Beweis

[0096] Die Vorhersage der Responsmatrix auf der Grundlage von  $n$  latenten Variablen, die beibehalten werden, ist folgendermaßen:

$$\hat{Y}_n = T_n \text{diag}\{b_n\} Q_n^T. \quad (A9)$$

[0097] Die  $t$ -Ergebnisse können direkt aus der ursprünglichen Prädiktormatrix berechnet werden, vergleiche Gleichung (A8):

$$T_n = X_n R_n. \quad (A10)$$

[0098] Die PLS-Regressionsmatrix kann daher folgendermaßen bestimmt werden:

$$B_{PLS}^{(n)} = R_n \text{diag}\{b_n\} Q_n^T. \quad (A11)$$

[0099] Schließlich stellt die vorherige Multiplikation mit  $P_n^T$  die erforderliche Gleichheit zur Verfügung (siehe Anhang 3):

$$P_n^T B_{PLS}^{(n)} = \text{diag}\{b_n\} Q_n^T. \quad (A12)$$

#### A.3 Für Gleichung (9) erforderlicher Beweis

[0100] Die Definition der Matrix  $R_n$ , vergleiche Gleichung (A8), kann dazu verwendet werden, zu beweisen, dass gilt:

$$R_n^T P_n = P_n^T R_n = \mathcal{I}_{n \times n}, \quad \mathcal{I}_{n \times n}$$

wobei  $\mathcal{I}_{n \times n}$  eine Einheitsmatrix  $n$  bezeichnet. Die Elemente des Matrixprodukts sind folgendermaßen definiert:

$$p_i^T r_j = p_i^T [\mathcal{I} - w_1 p_1^T] \dots [\mathcal{I} - w_{j-1} p_{j-1}^T] w_j; \quad 1 \leq i, j \leq n \quad (A13)$$

[0101] Wenn  $i$  größer ist als  $j$ , ist Gleichung (A13) gleich Null, da Skalarprodukte zwischen  $p_j^T$  und  $w_m$  auftreten, mit  $1 \leq m \leq j$ , was jeweils gleich Null ist [Hoskuldsson, 1988]. Gemäß Gleichung (A13) können, falls  $i$  kleiner als  $j$  ist, die Faktoren des Matrixproduktes bis zum  $i$ -ten Faktor reduziert werden, was führt zu:

$$p_i^T r_j = (p_i^T - p_i^T w_1 p_1^T) \dots [\mathcal{I} - w_{j-1} p_{j-1}^T] w_j; \quad p_i^T w_i = \frac{t_i^T X_{i-1} w_i}{t_i^T t_i} = 1$$

$$p_i^T r_i = 0 \quad (A14)$$

[0102] Gleichung (A14) zeigt, dass der transponierte Vektor in dem Term auf der rechten Seite gleich Null ist, und daher das gesamte Produkt gleich Null ist. Weiterhin führt die Tatsache, dass  $p_i^T w_i = 1$  ist, auch dazu, dass  $p_i^T r_i = 1$  ist. Zusammenfassend wird aus den Gleichungen (A13) und (A14) deutlich, dass das Matrixprodukt  $R_n^T P_n$  gleich einer Einheitsmatrix  $n \cdot n$  ist.

#### A.4 Orthogonalität beider verallgemeinerter Ergebnisse

[0103] Um zu beweisen, dass beide verallgemeinerten Ergebnisse nicht orthogonal sind, werden die Kovarianzmatrizen für beide Ergebnisse, einschließlich von  $n$  LVs, untersucht. Die Kovarianzmatrix der verallgemeinerten  $t$ -Ergebnisse ist in Gleichung (A15) angegeben:

$$\begin{aligned}
S_{T^*T^*}^{(n)} &= \frac{1}{k-1} [T_n + E_n^*]^T [T_n + E_n^*] = S_{TT}^{(n)} + S_{TE^*}^{(n)} + S_{TE^*}^{(n)T} + S_{E^*E^*}^{(n)} \\
S_{TE^*}^{(n)} &= \frac{1}{k-1} R_n^T X^T [E_n : F_n] C_{PLS}^{(n)} = R_n^T S_{XY}^{(n)} [\mathcal{I} - R_n P_n^T] [B_{PLS}^{(n)T} B_{PLS}^{(n)} + \mathcal{I}]^{-1} R_n \\
S_{E^*E^*}^{(n)} &= R_n^T [B_{PLS}^{(n)T} B_{PLS}^{(n)} + \mathcal{I}]^{-1} [B_{PLS}^{(n)T} S_{E_n E_n}^{(n)} B_{PLS}^{(n)} + S_{E_n E_n}^{(n)T} B_{PLS}^{(n)} + B_{PLS}^{(n)T} S_{F_n F_n}^{(n)T} + S_{F_n F_n}^{(n)}] [B_{PLS}^{(n)T} B_{PLS}^{(n)} + \mathcal{I}]^{-1} R_n
\end{aligned} \tag{A15}$$

[0104] In Gleichung (A15) repräsentieren

$$S_{T^*T^*}^{(n)}, S_{TT}^{(n)}, S_{E^*E^*}^{(n)}, S_{XX}^{(n)}, S_{F_n E_n}^{(n)} \text{ und } S_{F_n F_n}^{(n)} \text{ die}$$

Kovarianzmatrix der verallgemeinerten t-Ergebnisse, die Standard-t-Ergebnisse, die verallgemeinerten restlichen Ergebnisse, die Prädiktorvariablen, den Vorhersagefehler der Responsvariablen, und die Residuen der Prädiktorvariablen. Weiterhin bezeichnet  $s_{TE^*}^{(n)}$  die Kreuzkovarianzmatrix der Standard-t-Ergebnisse und der verallgemeinerten restlichen Ergebnisse, und ist  $s_{E_n E_n}^{(n)}$  die Kreuzkovarianzmatrix des Vorhersagefehlers der Responsmatrix und der Residuen der Prädiktorvariablen. Es wird darauf hingewiesen, dass nur die Kovarianzmatrix  $s_{TT}^{(n)}$  diagonal ist, da die Standard-t-Ergebnisse gegenseitig orthogonal sind [Hoskuldsson, 1988]. Aus Gleichung (A15) kann man schließen, dass die Matrizen  $s_{E^*E^*}^{(n)}$  und  $s_{TE^*}^{(n)}$  im Allgemeinen nicht diagonal sind, selbst unter der Annahme, dass die Spalten der Residuenmatrizen  $E_n$  und  $F_n$  und ebenso die Prädiktorvariablen  $X$  aus Signalen in Form weißen Rauschens bestehen. Im theoretischen Fall, in welchem die Anzahl an Responsvariablen gleich der Anzahl manipulierter Variablen ist, und der Prozess entkoppelt ist, sind jedoch die Kovarianzmatrix  $s_{E^*E^*}^{(n)}$  und die Kreuzkovarianzmatrix  $s_{TE^*}^{(n)}$  vom Diagonaltyp. Daher kann die Kovarianzmatrix  $s_{T^*T^*}^{(n)}$  nicht vom Diagonaltyp sein, wenn eine Anzahl  $n$  an LVs zurückbehalten wird. Wenn alle  $M$  (siehe Gleichung 1) LVs beibehalten werden, dann reduziert sich Gleichung (A15) auf:

$$S_{T^*T^*}^{(n)} = S_{TT}^{(n)} + R_M^T [B_{PLS}^{(M)T} B_{PLS}^{(M)} + \mathcal{I}]^{-1} B_{PLS}^{(M)T} S_{E_n E_n}^{(n)} B_{PLS}^{(M)} [B_{PLS}^{(M)T} B_{PLS}^{(M)} + \mathcal{I}]^{-1} R_M. \tag{A16}$$

[0105] Auf der Grundlage der Annahme, dass  $E_n$  aus Signalen in Form weißen Rauschens besteht, ist die Kovarianzmatrix  $S_{E^*E^*}$  offensichtlich vom Diagonaltyp. Infolge der vorherigen und nachfolgenden Multiplikation mit im Allgemeinen nicht-diagonalen Matrizen ergibt sich jedoch das Ergebnis, dass  $s_{E^*E^*}^{(n)}$  und daher  $s_{T^*T^*}^{(n)}$  nur symmetrisch sind. Die Interpretation der Gleichung (A15) und (A16) hat gezeigt, dass weder die Kovarianzmatrix  $s_{E^*E^*}^{(n)}$  noch  $s_{T^*T^*}^{(n)}$  diagonal ist. Daher sind beide verallgemeinerten Ergebnistypen im Allgemeinen nicht orthogonal.

#### LITERATUR

- Box G. E. P., Hunter W., Hunter J. S., 1978, Statistics of Experimenters, Wiley, New York
- Chen Q., Wynne R., Sandoz D. J., 2000. The Application of Principal Component Analysis and Kernel Density Estimation to Enhance Process Monitoring, Journal of Control Engineering Practise, to be published in April
- Dunia R., Qin S. J., Edgar T. F., McAvoy T. J., 1996, Identification of Faulty Sensors Using Principals Component Analysis, Process Systems Engineering, Vol. 42, No. 10, pp. 2797–2812
- McFarlane R. C., Reineman R. C., Bartee J. F., Georgakis C., 1993. Dynamic Simulator for a Model IV Fluid Catalytic Cracking Unit, Computers in Chemical Engineering, Vol. 17, No. 3, pp. 275–300
- Geladi P., 1986, PLS Tutorial, Analytica Chimica Acta, Vol. 185: pp. 1–17
- P. Geladi, 1988. Notes on the History and Nature of Partial Least Squares (PLS) Modelling, Journal of Chemometrics, Vol. 2, pp. 231–246
- Golub G. H., Van Loan C. F., 1996, Matrix Computations, Third Edition, John Hopkins University Press, Baltimore
- Hoskuldsson A., 1988. PLS Regression Methods, Journal of Chemometrics, Vol. 2: 211–228
- Jackson J. E., 1991, A User's Guide to Principal Components, Wiley, New York
- De Jong S., 1993, SIMPLS, an Alternative Approach to Partial Least Squares Regression, Chemometrics and intelligent Laboratory Systems, Vol. 18, pp. 251–263
- Kosanovich K. A., Piovoso M. J., 1991, Process Data Analysis Using Multivariate Statistical Methods, American Control Conference, Boston, Massachusetts
- Kourti T., MacGregor J. F., 1995, Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods, Chemometrics and intelligent Laboratory Systems, Vol. 28, pp. 3–21
- Kresta J. V., MacGregor J. F., Marlin T. E., 1991, Multivariate Statistical Monitoring of Process Operating Performance, The Canadian Journal of Chemical Engineering, Vol. 69, pp. 35–47
- Lindgren F., Geladi P., Wold S., 1993, The Kernel Algorithm for PLS, Journal of Chemometrics, Vol. 7, pp. 45–59
- MacGregor J. F., Marlin T. E., Kresta J., Skagerberg B., 1991, Multivariate Statistical Methods in Process Analysis and Control. AIChE Symposium Proceedings of the Fourth International Conference on Chemical Process Control, AIChE Publ. No. P-67, New York, pp. 79–99
- MacGregor J. F., Kourti T., 1995, Statistical Process Control of Multivariate Processes. Control Engineering Practise, Vol. 3, No. 3, pp. 403–414



17. Morud T. E., 1996, Multivariate Statistical Process Control; Example from the Chemical Process Industry, Journal of Chemometrics, Vol. 10, pp. 669-675
18. Nimmo L., 1995, Adequately Address Abnormal Situation Operations. Chemical Engineering Progress, Vol. 91, No. 1, pp. 36-45
19. Nomikos P., MacGregor J. F., 1994, Multi-Way Partial Least Squares for Monitoring Batch Processes, First International Chemometrics Internet Conference, September 1994
20. Nomikos, P., MacGregor. J. F., 1995. Multivariate SPC Charts for Monitoring Batch Processes, Technometrics, Vol. 37, No. 1, pp. 41-59
21. Manne R., 1987, Analysis of Two Partial-Least Squares Algorithms for Multivariate Calibration, Chemometrics and intelligent Laboratory Systems, Vol. 2, pp. 283-290
22. Piovoso M. J., Kosanovich K. A., 1992, Process Data Chemometrics, IEEE Transactions on Instrumentation and Measurement, Vol. 41, No. 2, pp. 262-268
23. Wise B. M., Gallagher N. B., MacGregor J. F., 1996. The Process Chemometrics Approach to Process Monitoring and Fault Detection, Journal of Process Control, Vol. 6. No. 6, pp. 329-348
24. H. Wold, 1966a, in Research Papers in Statistics, ed. by F. David p. 411, Wiley, New York
25. H. Wold, 1966b, in Multivariate Analysis, ed. by P. Krishnaiah, p. 391, Academic Press, New York
26. Wold S., 1978, Cross Validatory Estimation of the Number of Components in Factor and Principal Component Models, Technometrics. Vol. 20, No. 4, pp. 397-405

## 20 Patentansprüche

1. Verfahren zum Konstruieren/Konfigurieren eines mehrdimensionalen Monitors für statistische Prozesse durch eine Vorgehensweise mit partiellen kleinsten Fehlerquadraten, mit einer Konstruktion der Prozessprädiktor- und Responsmatrizen aus Bezugsdaten, wobei die Prädiktormatrix aus Signalen der manipulierten und gemessenen Störungs- oder Ursachenvariablen des Prozesses (Prädiktorvariablen) besteht, und die Responsmatrix aus den gesteuerten oder verursachten Variablen des Prozesses (Responsvariablen) besteht, Zerlegen der Prädiktor- und Responsmatrizen in Komponentenmatrizen des Rangs Eins, wobei jede der Matrizen aus einem Vektorprodukt besteht, bei welchem ein Vektor (der Ergebnisvektor) die Variation beschreibt, und der andere (der Belastungsvektor) den Beitrag des Ergebnisvektors zu der Prädiktor- oder Responsmatrix, wobei die Zerlegung durch die Erzeugung einer parametrischen Regressionsmatrix auf der Grundlage von Iterationen der Zerlegung der Prädiktor- und Responsmatrizen durchgeführt wird, **gekennzeichnet durch** die Erzeugung eines ersten verallgemeinerten Ergebnisvektors, der irgendeine signifikante Variation des Prozesses einschließlich Variationen der Prädiktor- und Responsvariablen beschreibt, und eines zweiten verallgemeinerten Ergebnisvektors, der den Vorhersagefehler des Modells der partiellen kleinsten Fehlerquadrate und Residuen der Prädiktormatrix repräsentiert.
2. Verfahren zum Konstruieren/Konfigurieren eines mehrdimensionalen Prozessmonitors nach Anspruch 1, bei welchem die verallgemeinerten Ergebnisse durch Konstruktion einer vergrößerten Matrix berechnet werden, die hier mit Z bezeichnet ist, und folgende Form aufweist:

$$Z = [Y : X],$$

wobei X die Prädiktormatrix und Y die Responsmatrix ist, und eine Ergebnismatrix  $T_n = T_n - E_n \cdot$  konstruiert wird, bei welcher  $T_n \cdot$  und  $E_n \cdot$  folgende allgemeine Form aufweisen:

$$T_n \cdot = [Y : X] [B_{PLS}^{(n)} : I]^T R_n$$

$$E_n \cdot = [E_n : F_n] [B_{PLS}^{(n)} : I]^T R_n$$

die Spalten der Matrix  $T_n \cdot$  die verallgemeinerten t-Ergebnisse zur Verfügung stellen, und die Spalten der Matrix  $E_n \cdot$  die verallgemeinerten restlichen Ergebnisse, wobei eine Einheitsmatrix  $M \times M$  bezeichnet, und

$$B_{PLS}^{(n)}$$

die PLS-Regressionsmatrix ist.

3. Mehrdimensionaler Monitor für statistische Prozesse, der durch das Verfahren von Anspruch 1 oder Anspruch 2 konstruiert/konfiguriert wurde, und so ausgebildet ist, dass er ein anomales Prozessverhalten durch Analyse der Residuen der Responsvariablen identifiziert.

4. Verfahren zur Überwachung eines Prozesses, welches das Konfigurieren eines mehrdimensionalen Monitors für statistische Prozesse durch das Verfahren nach Anspruch 1 oder Anspruch 2 umfasst, und Identifizieren eines anomalen Prozessverhaltens, zumindest teilweise, durch Analyse der Residuen der Responsvariablen.

Hierzu 13 Seite(n) Zeichnungen

- Leerseite -

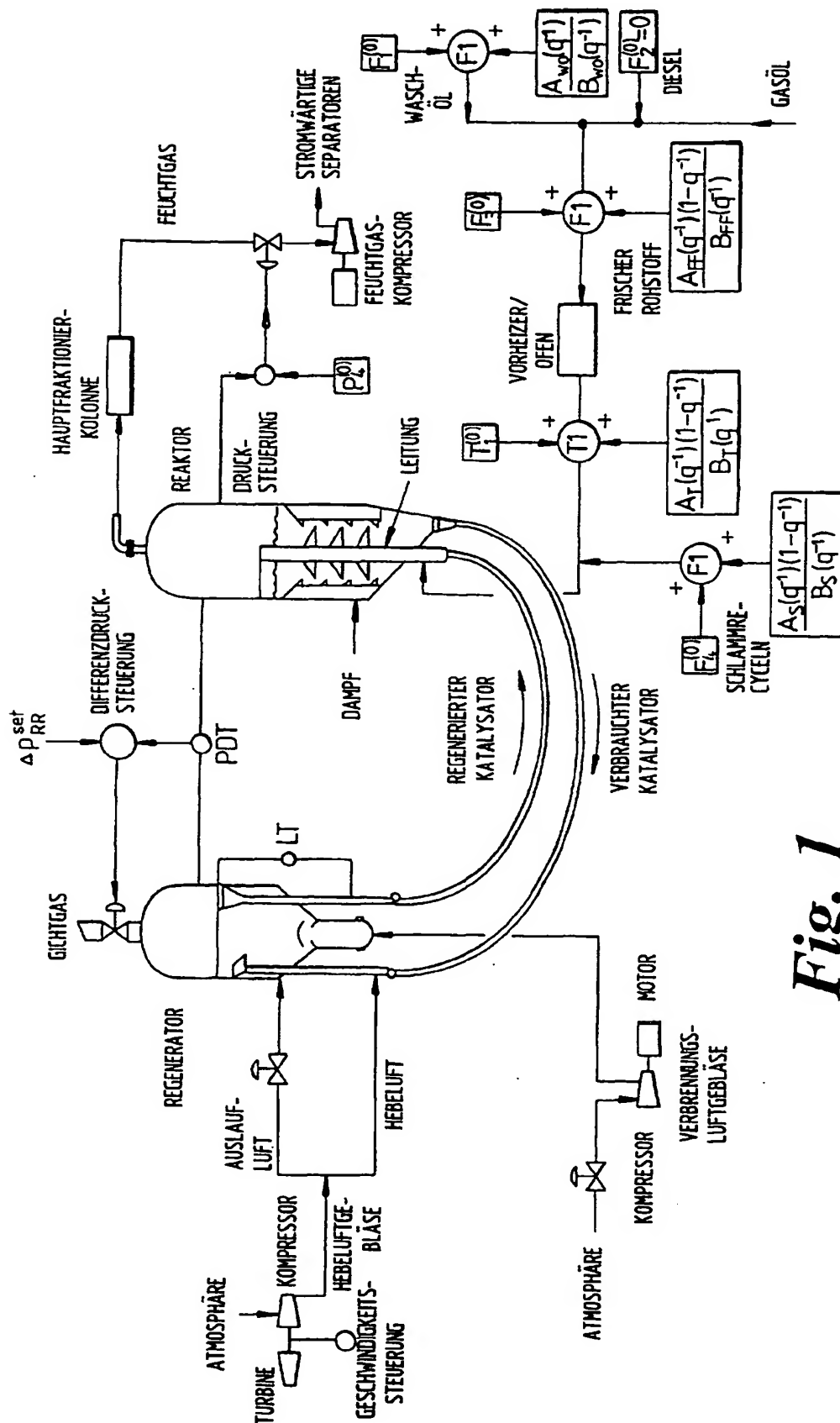


Fig. 1

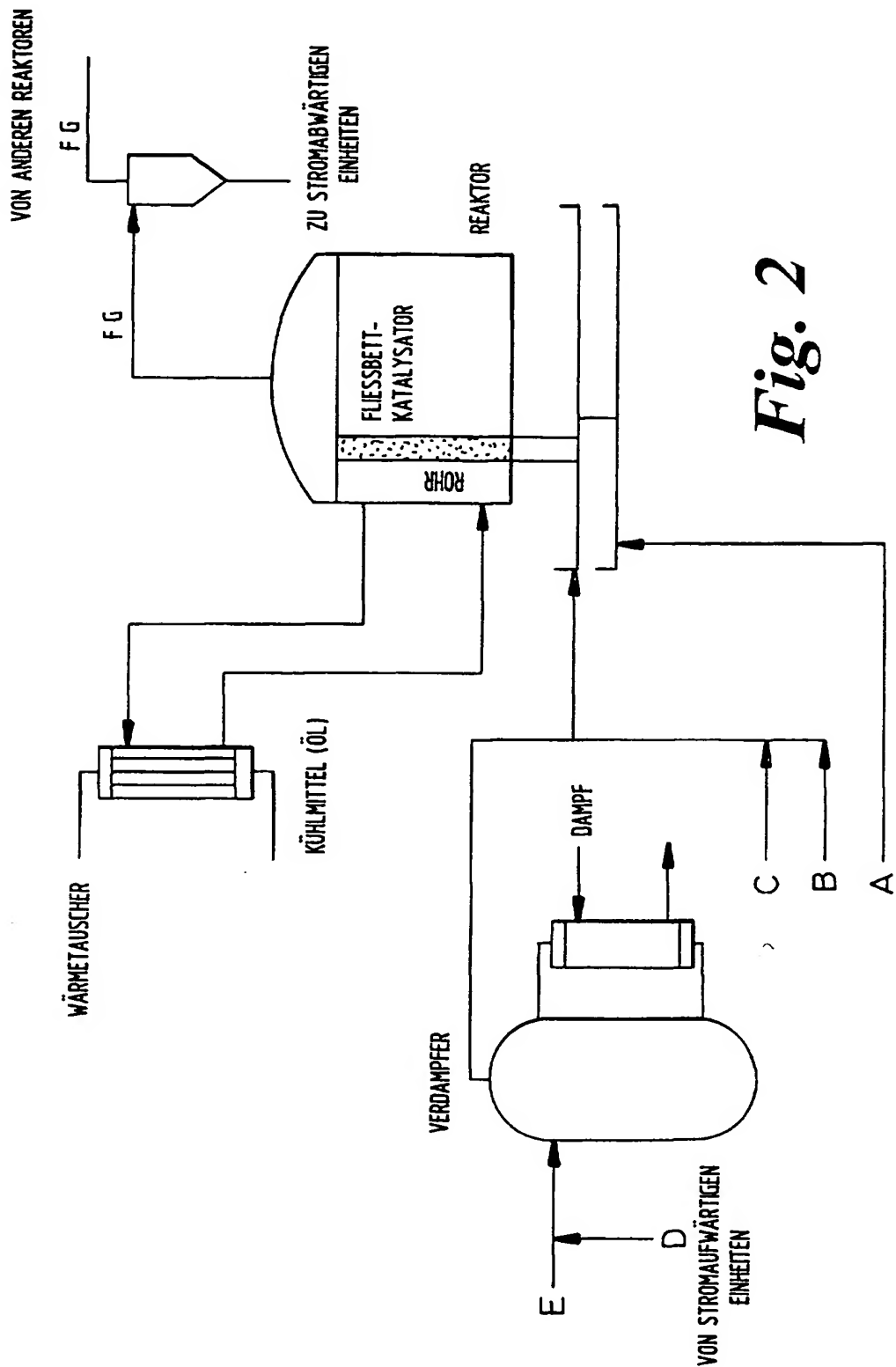
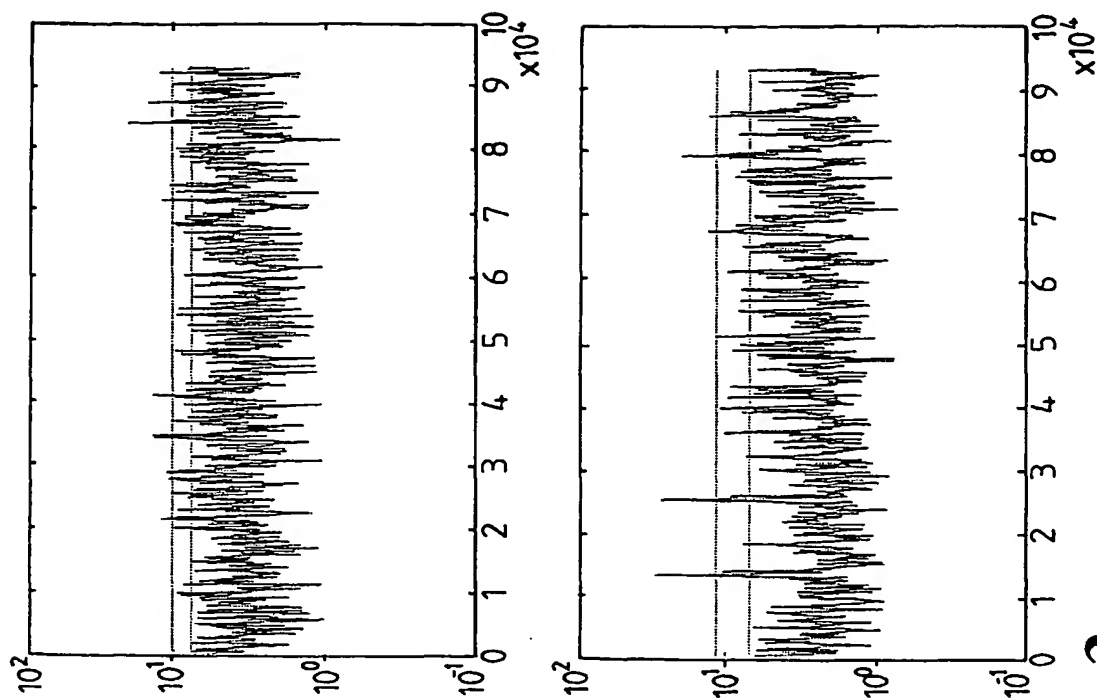
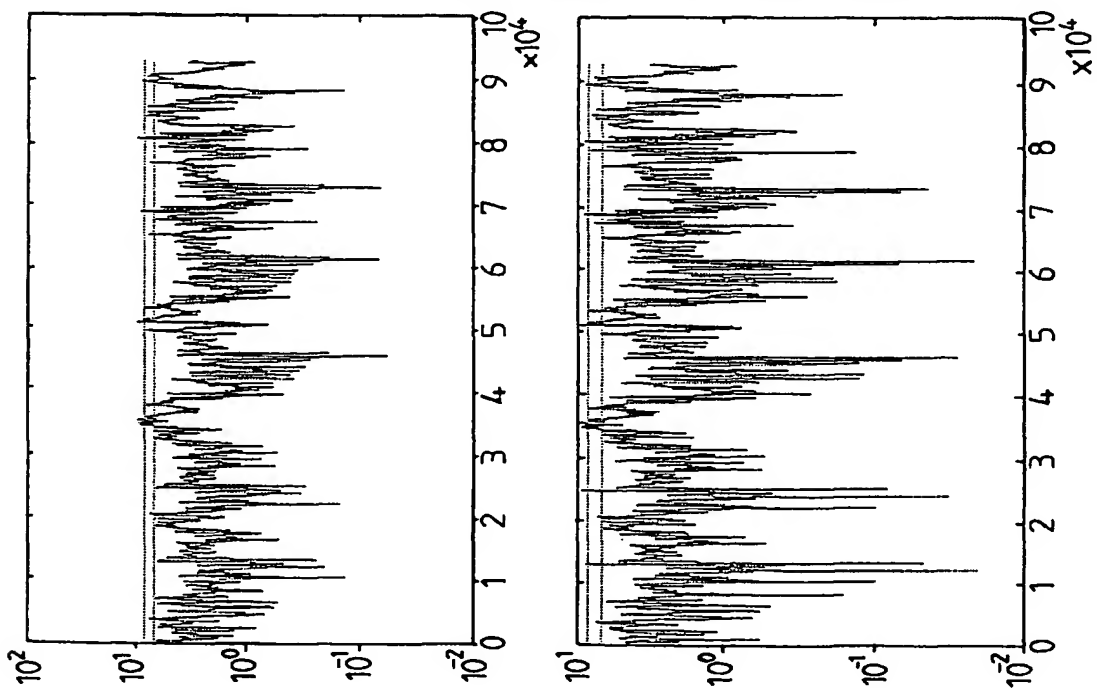


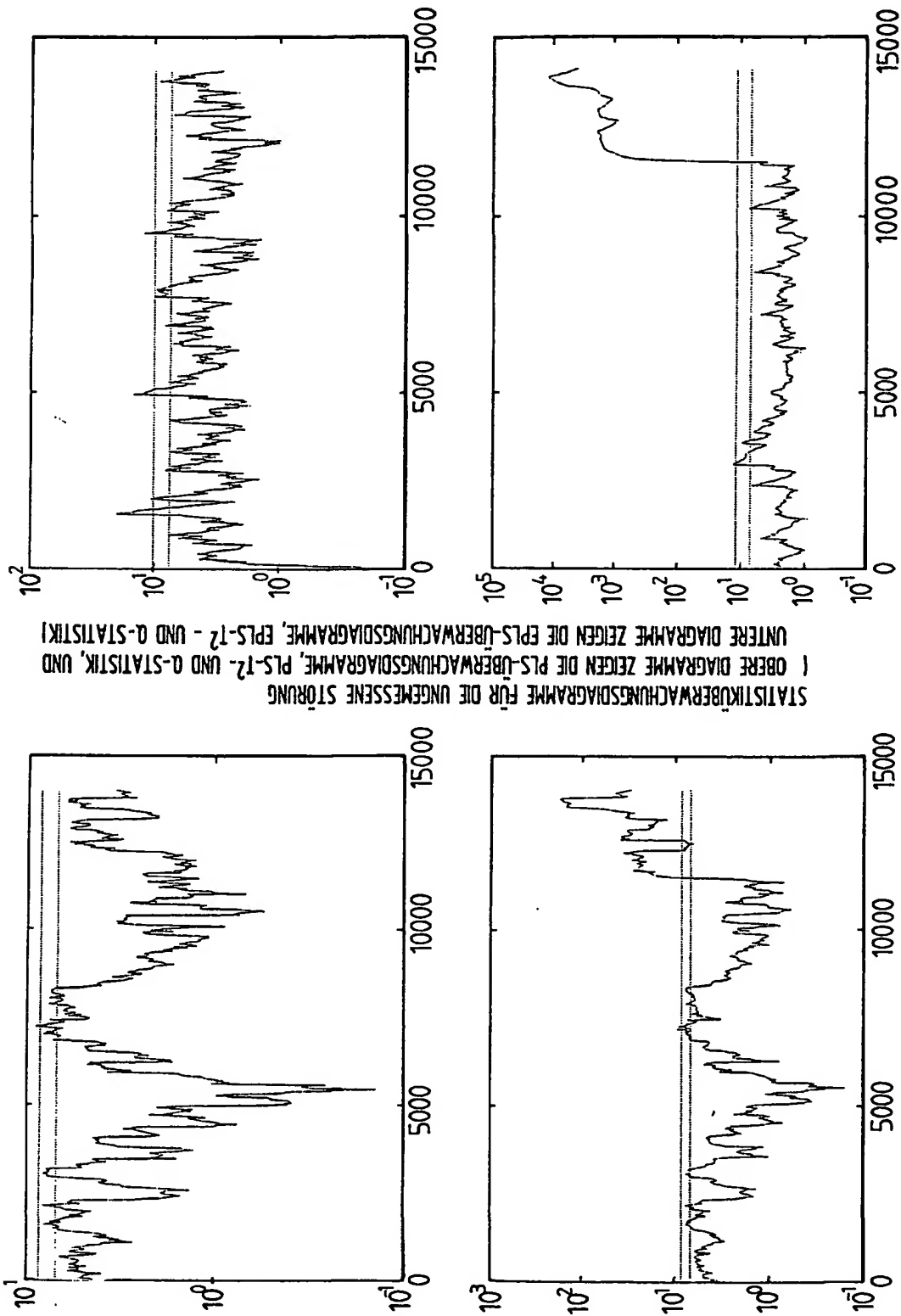
Fig. 2



STATISTIKÜBERWACHUNGSDIAGRAMME FÜR NORMALE BETRIEBSDATEN  
(OBERE DIAGRAMME ZEIGEN DIE PLS-ÜBERWACHUNGSDIAGRAMME, PLS-T2- UND Q-STATISTIK, UND  
UNTERE DIAGRAMME ZEIGEN DIE EPLS-ÜBERWACHUNGSDIAGRAMME, EPLS-T2- UND Q-STATISTIK)

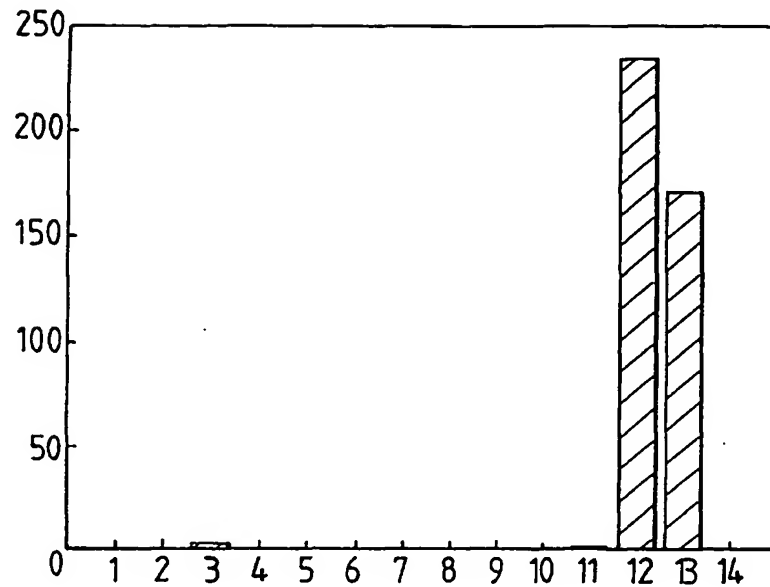
*Fig. 3*



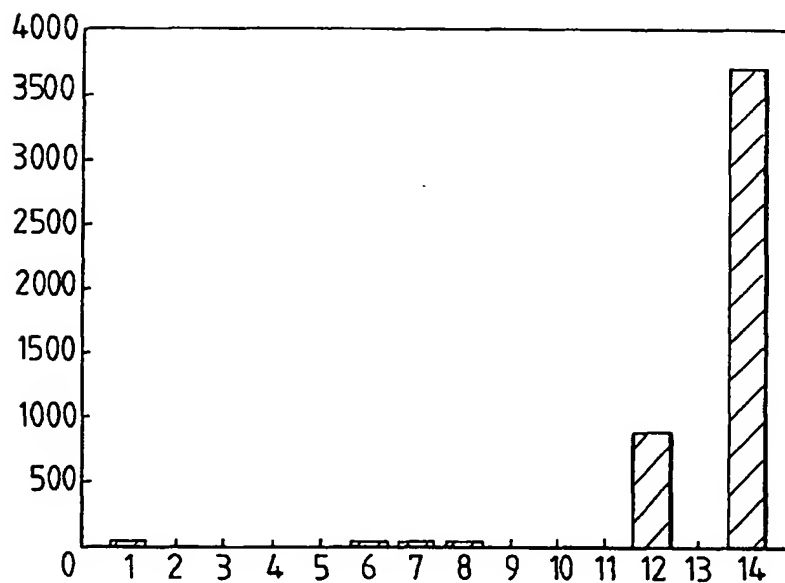


STATISTIKÜBERWACHUNGSDIAGRAMME FÜR DIE UNGEMESSENE STÖRUNG  
 ( OBERE DIAGRAMME ZEIGEN DIE PLS-ÜBERWACHUNGSDIAGRAMME, PLS-T<sup>2</sup>- UND Q-STATISTIK, UND  
 UNTERE DIAGRAMME ZEIGEN DIE EPLS-ÜBERWACHUNGSDIAGRAMME, EPLS-T<sup>2</sup>- UND Q-STATISTIK)

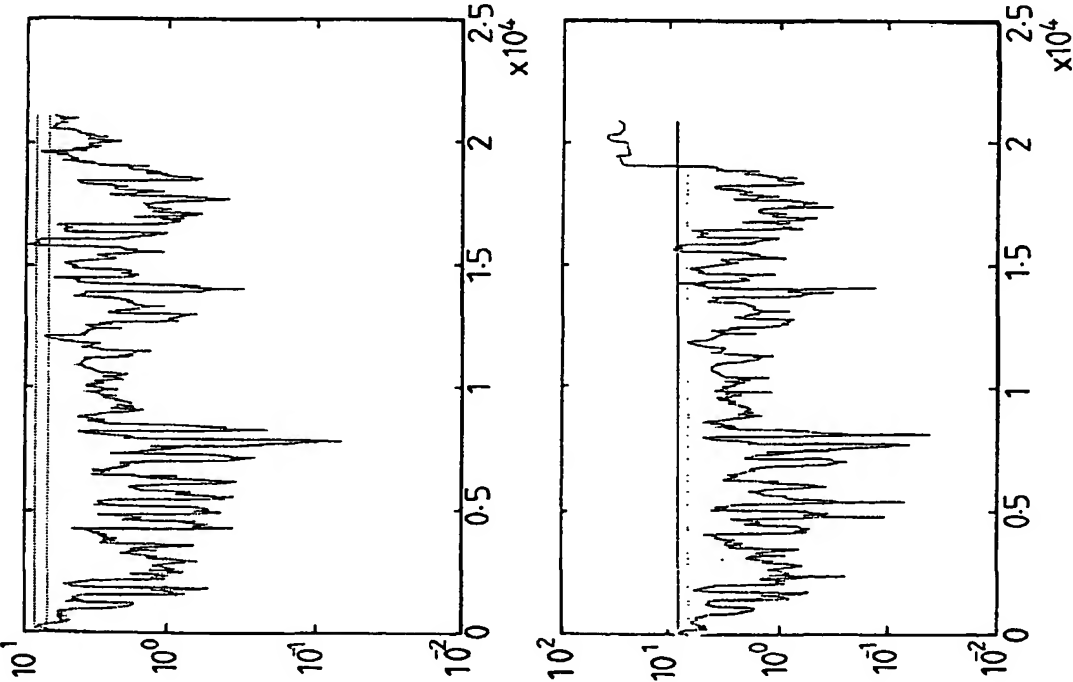
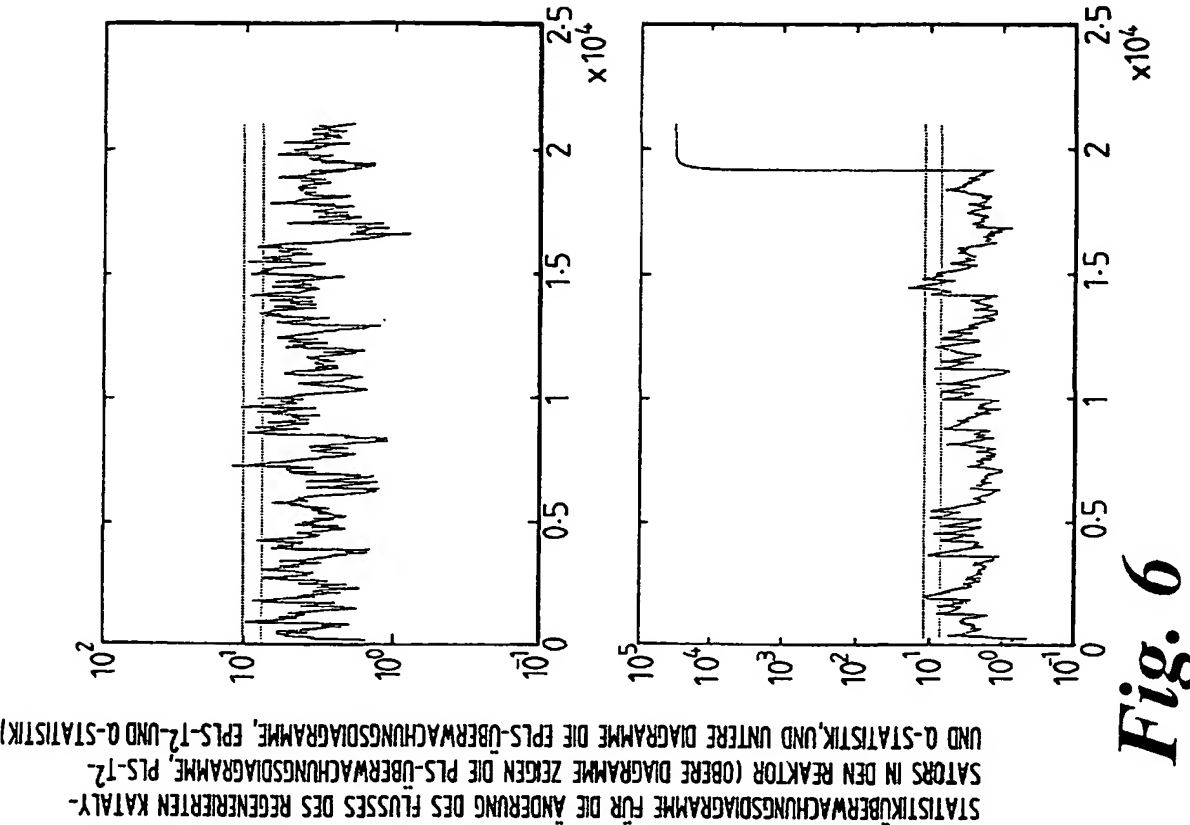
Fig. 4

**Fig. 5**

FEHLERBEITRAGSDIAGRAMM FÜR ZEITPUNKT 11460 min. DIE O<sub>2</sub> - UND CO-KONZENTRATION IN DEM GICHTGASFLUSS HABEN DEN GRÖSSTEN VORHERSAGEFEHLER.

**Fig. 7**

FEHLERBEITRAGSDIAGRAMM FÜR DIE ÄNDERUNG DES FLUSSES DES REGENERIERTEN KATALYSATORS IN DEN REAKTOR ZUM ZEITPUNKT 19357 min. DER STANDROHR-KATALYSATORPEGEL UND DIE O<sub>2</sub> - KONZENTRATION IM GICHTGAS WERDEN AM STÄRKSTEN BEEINFLUSST





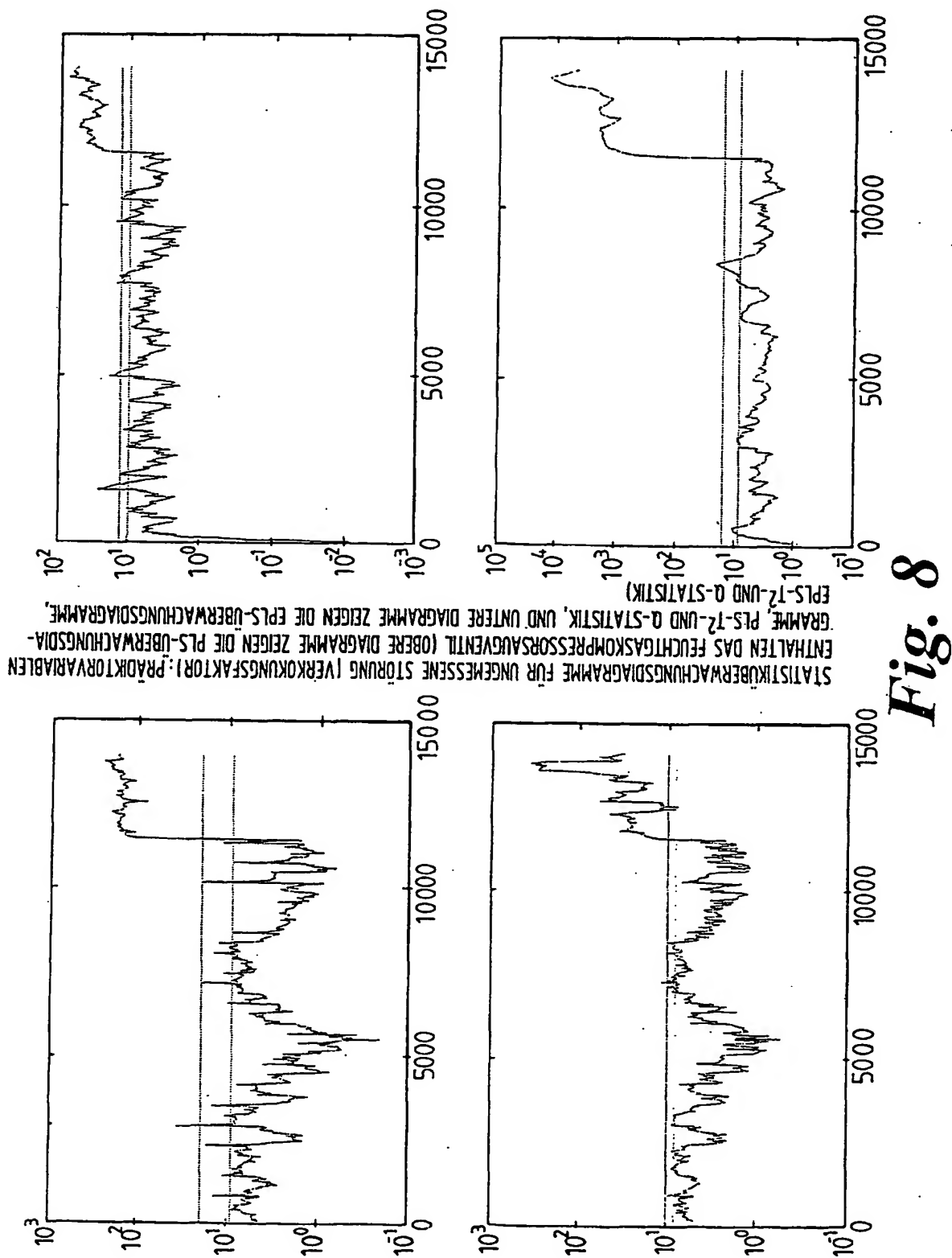


Fig. 8

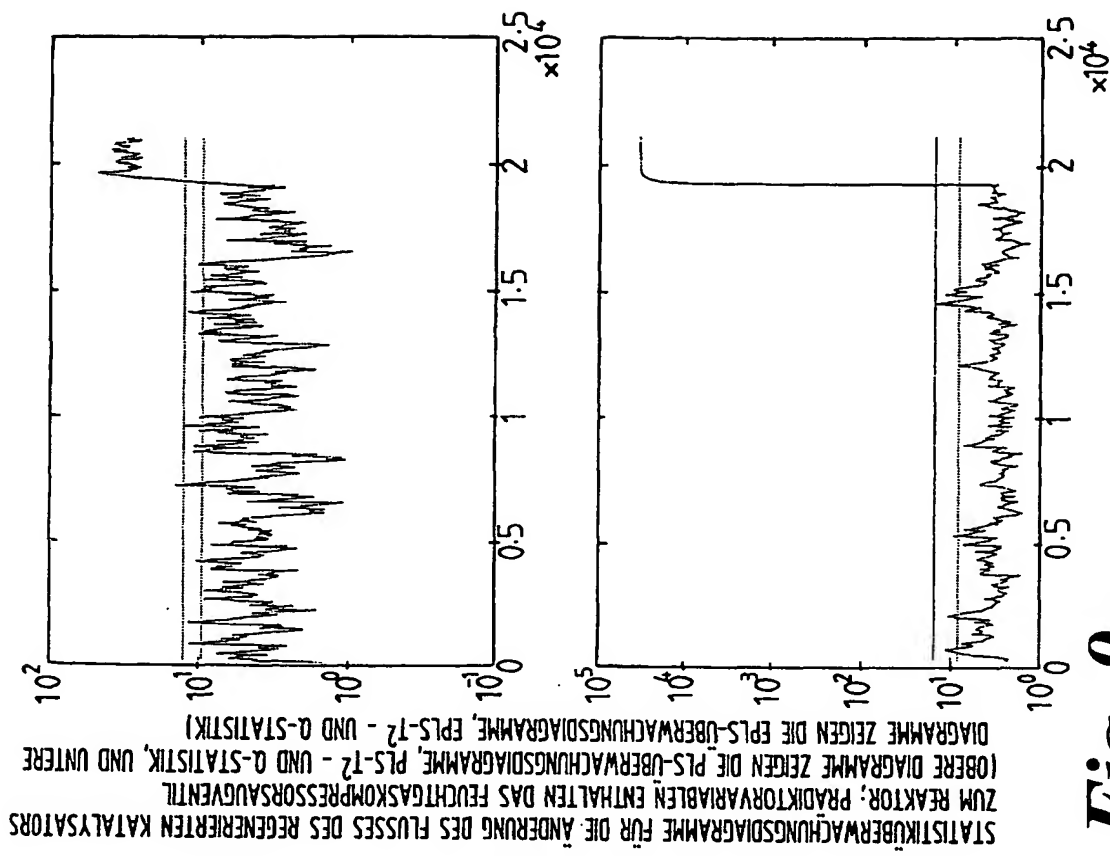
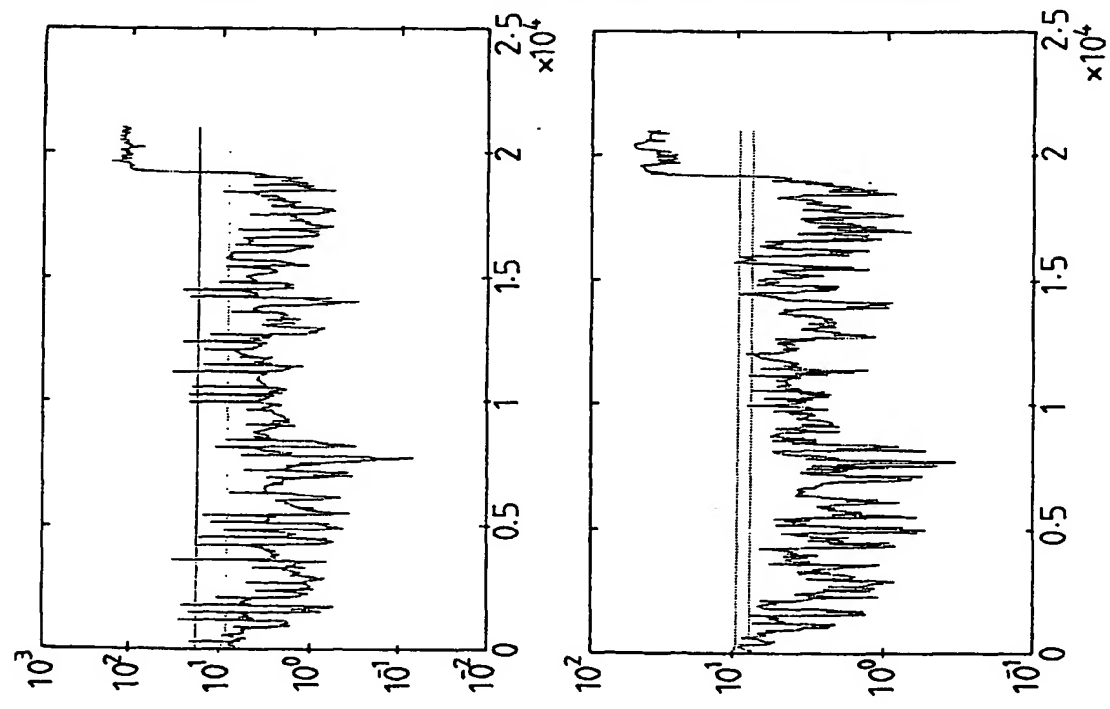
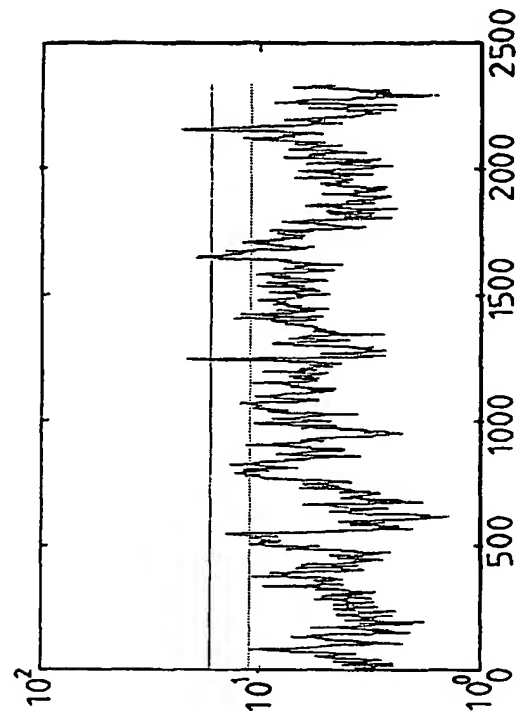
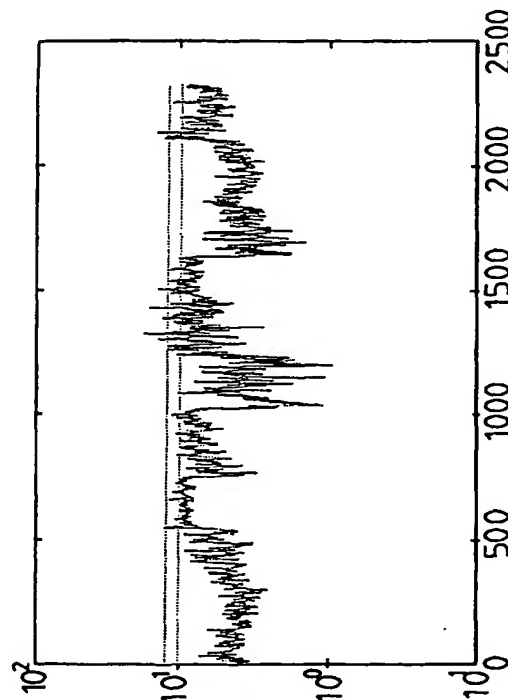
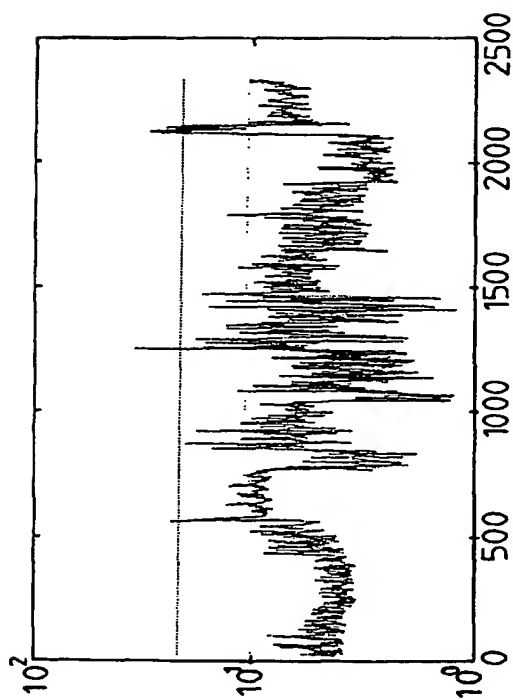


Fig. 9



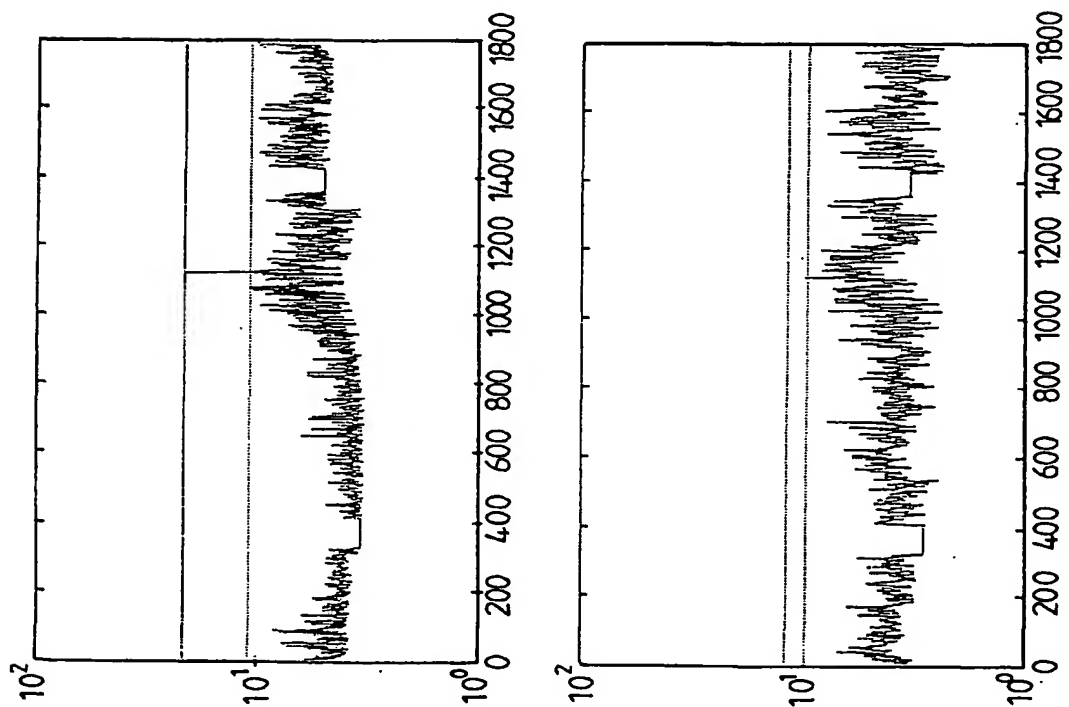
STATISTIKÜBERWACHUNGSDIAGRAMME FÜR NORMALBETRIEBSDATEN  
 (OBERES DIAGRAM ZEIGT DAS PLS-ÜBERWACHUNGSDIAGRAMM, PLS-T<sup>2</sup> - UND Q-STATISTIK,  
 UND UNTERE DIAGRAMME ZEIGEN DIE EPLS-ÜBERWACHUNGSDIAGRAMME, EPLS-T<sup>2</sup> - UND  
 Q-STATISTIK)



**Fig. 10**

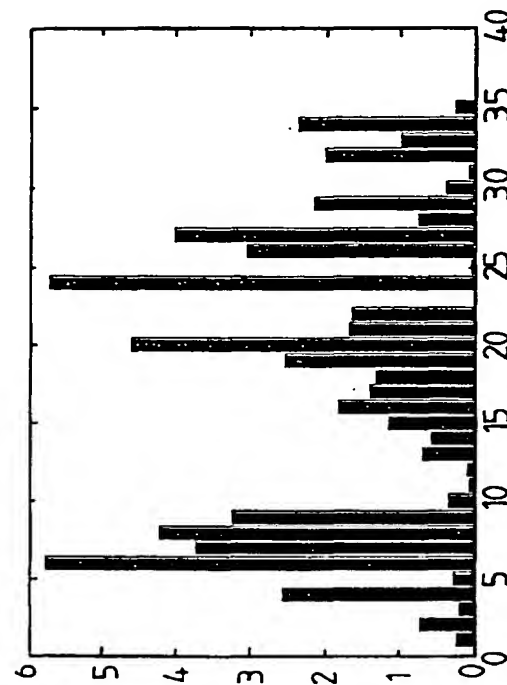
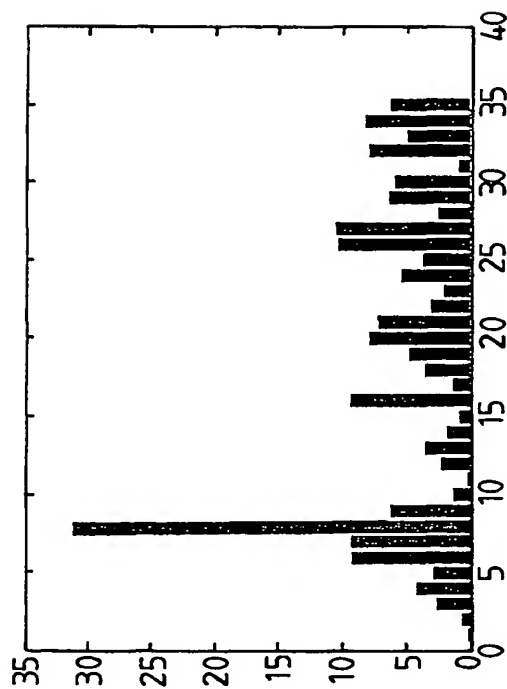
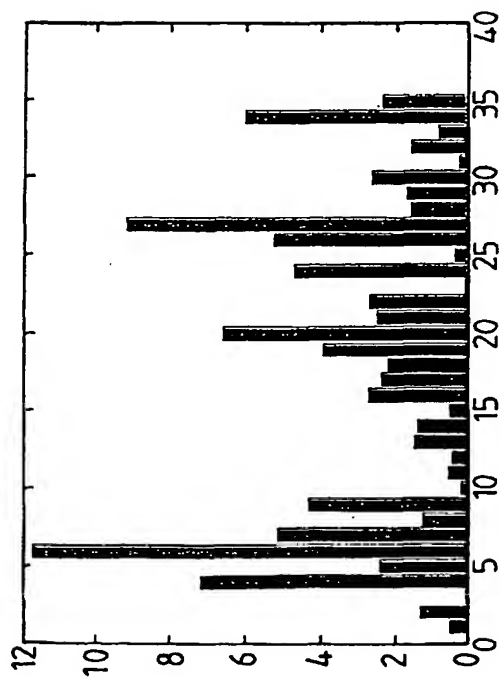
**Fig. 11**

STATISTIKÜBERWACHUNGSDIAGRAMME FÜR DIE UNGEMESSENE STÖRUNG  
(OBERES DIAGRAMM ZEIGT DAS PLS-ÜBERWACHUNGSDIAGRAMM, PLS-T<sup>2</sup>- UND Q-STATISTIK,  
UND UNTERE DIAGRAMME ZEIGEN DIE EPLS-ÜBERWACHUNGSDIAGRAMME, EPLS-T<sup>2</sup>-  
UND Q-STATISTIK)



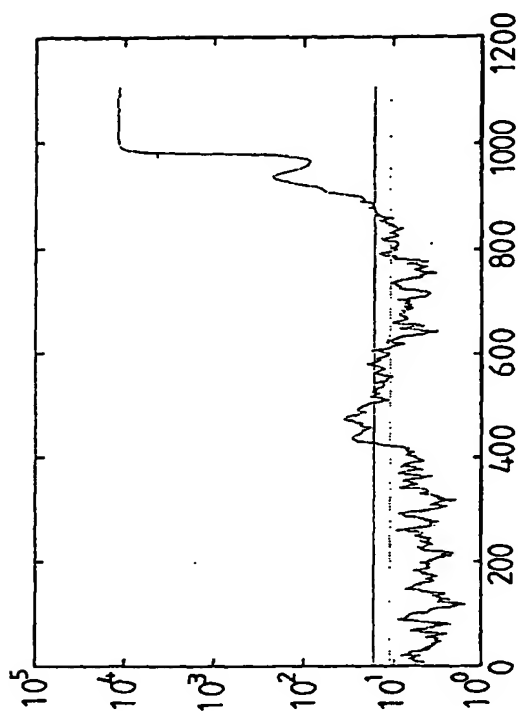
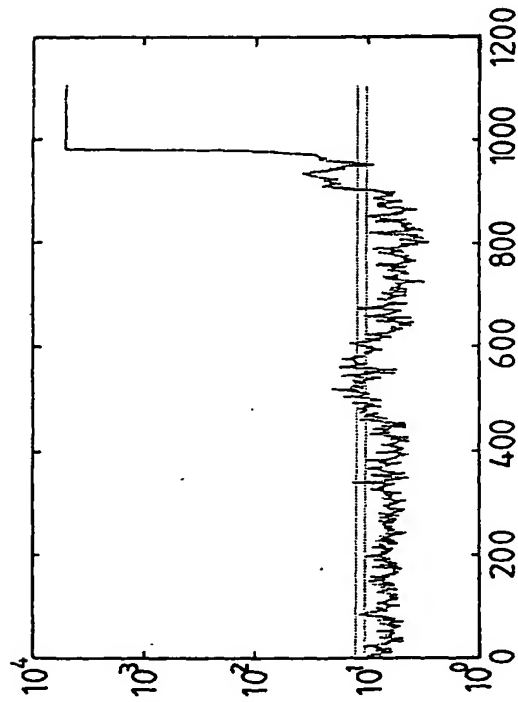
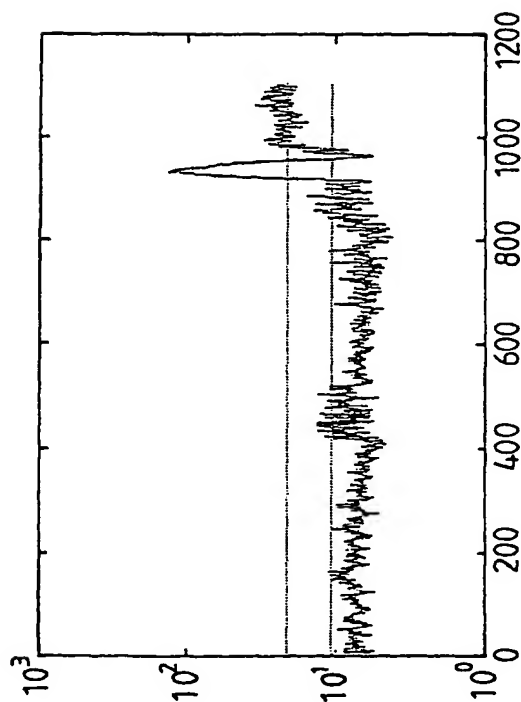
*Fig. 12*

EC-DIAGRAMME FÜR DAMFERDRUCKSCHLAG ZU DEN ZEITPUNKTEN  
1500 min (OBEN LINKS), 1501 min (UNTEN LINKS) UND 1502 min (UNTEN RECHTS)



**Fig. 13**

STATISTIKÜBERWACHUNGSDIAGRAMME FÜR ANORMALES VERHALTEN EINES DER RÖHRE  
(OBERESDIAGRAMM ZEIGT DAS PLS-ÜBERWACHUNGSDIAGRAMM, PLS- $T^2$ - UND Q-STATISTIK, UND  
UNTERE DIAGRAMME ZEIGEN DIE EPLS-ÜBERWACHUNGSDIAGRAMME, EPLS- $T^2$ - UND Q-STATISTIK)



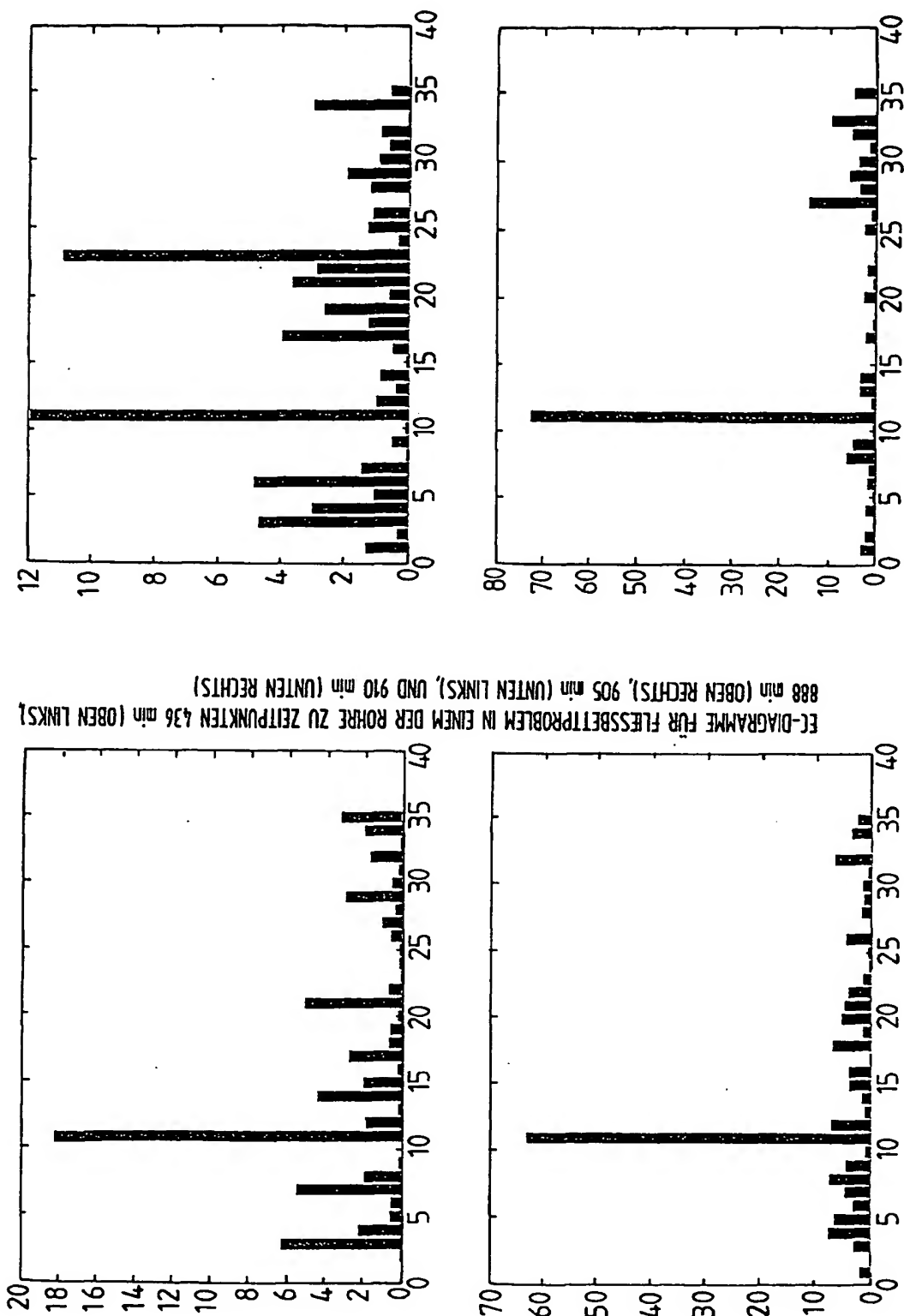


Fig. 14